

> 0 <
0| 0 IntelliGenetics
> 0 <

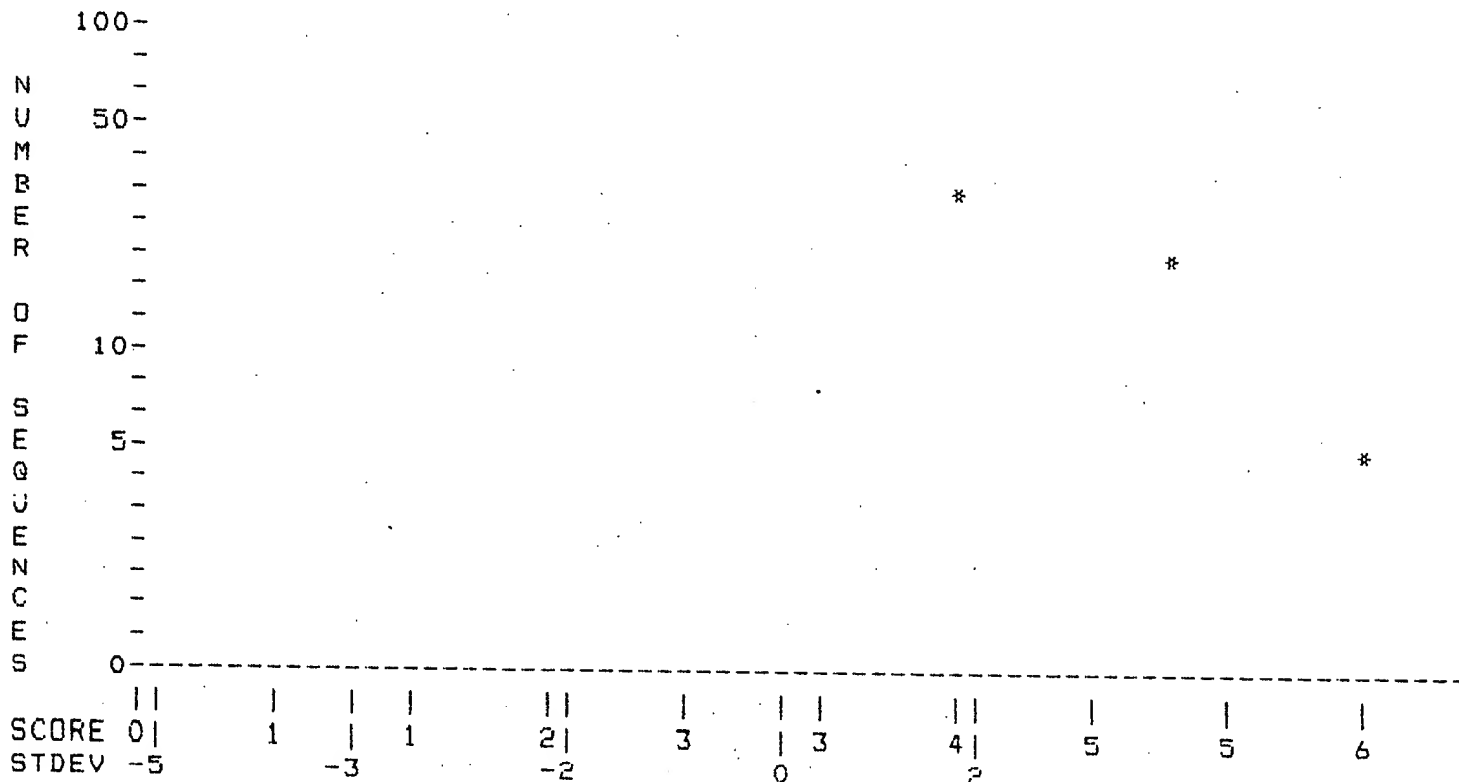
FastDB --Fast Pairwise Comparison of Sequences
Release 5.4

*Sequences numbered
as in table 4 1-19
(not by seq ID #)*

Results file celsa-1-align.res made by alexk on Thu 25 Feb 93 10:46:11-PST.

Query sequence being compared: CELSA-1 (1-6)
Number of sequences searched: 50
Number of scores above cutoff: 50

Results of the initial comparison of CELSA-1 (1-6) with:
File : celsa-1.pep



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	5	0.65

Times:	CPU	Total Elapsed
	00:00:00.03	00:00:01.00

Number of residues:	30314
Number of sequences searched:	50
Number of scores above cutoff:	50

Significance is calculated based on initial score.

4 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. GNWVNE	Genome polyprotein - Tick-bor	3414	6	6	3.09	0
2. POLG_TBEVW	GENOME POLYPROTEIN (CONTAINS:	3414	6	6	3.09	0
3. GNWVTB	Genome polyprotein - Tick-bor	3412	6	6	3.09	0
4. POLG_TBEVS	GENOME POLYPROTEIN (CONTAINS:	3412	6	6	3.09	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 1 standard deviation above mean ****						
5. A39246	*Carboxypeptidase A precursor	417	5	5	1.55	0
6. CBPC_HUMAN	MAST CELL CARBOXYPEPTIDASE A	417	5	5	1.55	0
7. S04912	*yopA protein - Yersinia ente	455	5	5	1.55	0
8. A39938	*Phosphotransferase system en	460	5	5	1.55	0
9. PT2S_BACSU	PHOSPHOTRANSFERASE ENZYME II,	460	5	5	1.55	0
10. WMBE71	UL13 protein - Human herpesvi	518	5	5	1.55	0
11. D30083	*Gene UL13 protein - Human he	518	5	5	1.55	0
12. KR2_HSV11	PROBABLE SERINE/THREONINE-PRO	518	5	5	1.55	0
13. DNBEV1	DNA-binding protein - Human h	1196	5	5	1.55	0
14. DNBEKS	DNA-binding protein - Human h	1196	5	5	1.55	0
15. DNBEHF	DNA-binding protein - Human h	1196	5	5	1.55	0
16. B30085	*Gene UL29 protein (major DNA	1196	5	5	1.55	0
17. DNBI_HSV11	MAJOR DNA-BINDING PROTEIN.	1196	5	5	1.55	0
18. DNBI_HSV1F	MAJOR DNA-BINDING PROTEIN.	1196	5	5	1.55	0
19. DNBI_HSV1K	MAJOR DNA-BINDING PROTEIN.	1196	5	5	1.55	0
20. S19926	*Hypothetical protein - Red c	1864	5	5	1.55	0

1. CELSA-1 (1-6)

```

GNWVNE      Genome polyprotein - Tick-borne encephalitis virus

ENTRY       GNWVNE      #Type Protein
TITLE       Genome polyprotein - Tick-borne encephalitis virus
             (subtype Western, strain Neudoerfl)
INCLUDES    capsid protein C\ membrane protein M\ envelope
             protein E\ nonstructural protein NS1\
             nonstructural protein NS2a\ nonstructural protein
             NS2b\ nonstructural protein NS3\ nonstructural
             protein NS4a\ nonstructural protein NS4b\
             nonstructural protein NS5
DATE        31-Dec-1989 #Sequence 30-Jun-1991 #Text 30-Sep-1992
PLACEMENT   2169.0      4.0      1.0      2.0      1.0
SOURCE      tick-borne encephalitis virus
ACCESSION   A31052\ A32596
REFERENCE
#Authors    Mandl C.W., Heinz F.X., Kunz C.
#Journal    Virology (1988) 166:197-205
#Title      Sequence of the structural proteins of tick-borne
             encephalitis virus (Western subtype) and
             comparative analysis with other flaviviruses.
#Reference-number A31052
#Accession   A31052
#Molecule-type genomic RNA
#Residues    1-779 <MAN1>
REFERENCE
#Authors    Mandl C.W., Heinz F.X., Stoeckl E., Kunz C.

```

#Reference-number A32596
 #Accession A32596
 #Molecule-type genomic RNA
 #Residues 767-3414 <MAN2>
 COMMENT This virus is a member of the family Flaviviridae.
 SUPERFAMILY #Name yellow fever virus genome polyprotein
 KEYWORDS capsid protein\ envelope protein\ glycoprotein\
 membrane protein\ nonstructural protein\
 polyprotein
 FEATURE
 2-116 #Protein capsid protein C <CPC>\
 117-280 #Protein membrane protein M precursor
 <MPP>\
 117-205 #Domain signal sequence <SIG>\
 206-280 #Protein membrane protein M <MPM>\
 246-264 #Domain transmembrane <TM1>\
 281-776 #Protein envelope protein E <EPE>\
 738-751 #Domain transmembrane <TM2>\
 777-1128 #Protein nonstructural protein NS1 <NS1>\
 1129-1358 #Protein nonstructural protein NS2a
 <N2A>\
 1359-1489 #Protein nonstructural protein NS2b
 <N2B>\
 1490-2110 #Protein nonstructural protein NS3 <NS3>\
 2111-2259 #Protein nonstructural protein NS4a
 <N4A>\
 2260-2511 #Protein nonstructural protein NS4b
 <N4B>\
 2512-3414 #Protein nonstructural protein NS5 <NS5>\
 144,434,641,753,861,
 983,999,1649,1988,
 2044,2447,2529,2686,
 2726 #Binding-site carbohydrate (Asn)
 (covalent) (predicted)

SUMMARY #Molecular-weight 378383 #Length 3414 #Checksum 1083

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:21:17-PST using FastDB

Initial Score = 6 Optimized Score = 6 Significance = 3.09
 Residue Identity = 100% Matches = 6 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

RRVTTASAAQRRGRVGRQDGRITDEYIYSGQDDDDSGLVQWKEAQILLDNITTLRGPVATFYGPEQDKMPEV
 1940 1950 1960 1970 1980 1990 2000 2010

X X

WHVAAN

|||||

AGHFRLTEEKRKHFRLHLLTHCDFTPLAWHVAANVSSVTDRTSWEGPDRDAVDEASGDLVTFRSPNGAERT
 2020 2030 2040 X 2050 2060 2070 2080

LRPVWKDARMFKEGRDIKEFVAYASGRRSFGDVLTMMSGVPELLRHRCVSALDVFYTLMHHEE
 2090 2100 2110 2120 2130 2140

2. CELSA-1 (1-6)

POLG_TBEVW GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CO

ID POLG_TBEVW STANDARD; PRT; 3414 AA.
 AC P14336;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)

DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
 DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
 DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
 DE RNA POLYMERASE (EC 2.7.7.48) (NS5)).
 OS TICK-BORNE ENCEPHALITIS VIRUS (WESTERN SUBTYPE) (TBEV).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE.
 RN [1]
 RP SEQUENCE OF 1-779 FROM N.A.
 RC STRAIN=NEUDOERFL;
 RM 88322870
 RA MANDL C.W., HEINZ F.X., KUNZ C.;
 RL VIROLOGY 166:197-205(1988).
 RN [2]
 RP SEQUENCE OF 767-3414 FROM N.A.
 RC STRAIN=NEUDOERFL;
 RM 90051080
 RA MANDL C.W., HEINZ F.X., STOECKL E., KUNZ C.;
 RL VIROLOGY 173:291-301(1989).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 DR EMBL; M21498; TOGTBESP.
 DR EMBL; M33668; TBEGNE.
 DR PIR; A31052; GNWVNE.
 KW POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
 KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
 KW NONSTRUCTURAL PROTEIN.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 2 112 CAPSID PROTEIN C.
 FT PROPEP 113 205
 FT CHAIN 206 280 ENVELOPE GLYCOPROTEIN M.
 FT CHAIN 281 776 MAJOR ENVELOPE PROTEIN E.
 FT CHAIN 777 1128 NONSTRUCTURAL PROTEIN NS1.
 FT CHAIN 1129 1358 NONSTRUCTURAL PROTEIN NS2A.
 FT CHAIN 1359 1489 NONSTRUCTURAL PROTEIN NS2B.
 FT CHAIN 1490 2110 HELICASE (NS3).
 FT CHAIN 2111 2259 NONSTRUCTURAL PROTEIN NS4A.
 FT CHAIN 2260 2511 NONSTRUCTURAL PROTEIN NS4B.
 FT CHAIN 2512 3414 RNA-DIRECTED RNA POLYMERASE (NS5).
 FT TRANSMEM 101 112 HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
 FT TRANSMEM 247 259 POTENTIAL.
 FT TRANSMEM 266 280 POTENTIAL.
 FT TRANSMEM 738 751 POTENTIAL.
 FT CARBOHYD 144 144 POTENTIAL.
 FT CARBOHYD 434 434 POTENTIAL.
 FT CARBOHYD 861 861 POTENTIAL.
 FT CARBOHYD 983 983 POTENTIAL.
 FT CARBOHYD 999 999 POTENTIAL.
 FT CARBOHYD 2447 2447 POTENTIAL.
 SQ SEQUENCE 3414 AA; 378379 MW; 2.052432E+07 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:21:16-PST using FastDB

 Initial Score = 6 Optimized Score = 6 Significance = 3.09
 Residue Identity = 100% Matches = 6 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

RRVTTASAAQRRGRVGRQDGRITDEYIYSGQDDDDSGLVQWKEAQILLDNITTLRGPVATFYGSPEQDKMPEV
 1940 1950 1960 1970 1980 1990 2000 2010

AGHFRLTEEKRKHFRLHLLTHCDFTPWLAWHVAANVSSVTDRSWTWEGPDRDAVDEASGDLVTFRSPNGAERT
 2020 2030 2040 X 2050 2060 2070 2080

LRPVWKDARMFKEGRDIKEFVAYASGRRSFGDVLGTMSGVPELLRHRCVSALDVFYTLMHHEE
 2090 2100 2110 2120 2130 2140

3. CELSA-1 (1-6)

GNWVTB Genome polyprotein - Tick-borne encephalitis virus

ENTRY GNWVTB #Type Protein
 TITLE Genome polyprotein - Tick-borne encephalitis virus
 (strain Sofjin)
 INCLUDES capsid protein C\ envelope protein prM\ envelope
 protein M\ major envelope protein E\ nonstructural
 protein NS1\ nonstructural protein NS2a\
 nonstructural protein NS2b\ nonstructural protein
 NS3\ nonstructural protein NS4a\ nonstructural
 protein NS4b\ nonstructural protein NS5
 DATE 31-Mar-1991 #Sequence 31-Mar-1991 #Text 30-Jun-1992
 PLACEMENT 2169.0 4.0 1.0 1.0 1.0
 SOURCE tick-borne encephalitis virus
 ACCESSION A33776\ S06414
 REFERENCE

#Authors Pletnev A.G., Yamshchikov V.F., Blinov V.M.
 #Journal Virology (1990) 174:250-263
 #Title Nucleotide sequence of the genome and complete amino
 acid sequence of the polyprotein of tick-borne
 encephalitis virus.

#Reference-number A33776
 #Accession A33776
 #Molecule-type genomic RNA
 #Residues 1-3412 <PLE>
 #Cross-reference GB:X07755

COMMENT This virus is a member of the family Flaviviridae.
 SUPERFAMILY #Name yellow fever virus genome polyprotein
 KEYWORDS capsid protein\ envelope protein\ glycoprotein\
 nonstructural protein\ polyprotein

FEATURE

2-112 #Protein capsid protein C <CPC>
 113-205 #Protein envelope protein prM <PRM>
 206-280 #Protein envelope protein M <PMM>
 281-776 #Protein major envelope protein E <PPE>
 777-1190 #Protein nonstructural protein NS1 <NS1>
 1191-1358 #Protein nonstructural protein NS2a
 <N2A>
 1359-1489 #Protein nonstructural protein NS2b
 <N2B>
 1490-2110 #Protein nonstructural protein NS3 <NS3>
 2111-2259 #Protein nonstructural protein NS4a
 <N4A>
 2260-2510 #Protein nonstructural protein NS4b
 <N4B>
 2511-3412 #Protein nonstructural protein NS5 <NS5>
 144,434,641,753,861,
 983,999,1228,1649,
 1988,2044,2052,2447,
 2466,2685,2725 #Binding-site carbohydrate (Asn)
 (covalent) (predicted)

SUMMARY #Molecular-weight 377979 #Length 3412 #Checksum 7007
 SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:21:17-PST using FastDB

Initial Score = 6 Optimized Score = 6 Significance = 3.09

RRVTTASAAQRRGRVGRQEGRTDEYIYSGQCDDDSGLVQWKEAQILLDNITTLRGFVATFYGPEQDKMPEV
 1940 1950 1960 1970 1980 1990 2000 2010

X X
 WHVAAN

|||||

AGHFRLTEEKRKHFRLHLLTHCDFTPWLAWHVAANVSSVTSRNWTWEGPEENTVDEANGDLVTRSPNGAERT
 2020 2030 2040 X 2050 2060 2070 2080

LRPVWRDARMFREGRDIREFVAYASGRRSFGDVLSGMSGVPELLRHRCVVSAMDVFYITLMHEE
 2090 2100 2110 2120 2130 2140

4. CELSA-1 (1-6)

POLG_TBEVS GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CO

ID POLG_TBEVS STANDARD; PRT; 3412 AA.
 AC P07720; P07721;
 DT 01-APR-1988 (REL. 07, CREATED)
 DT 01-MAY-1991 (REL. 18, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE GENOME POLYPROTEIN (CONTAINS: CAPSID PROTEIN C (CORE PROTEIN); MATRIX
 DE PROTEIN (ENVELOPE PROTEIN M); MAJOR ENVELOPE PROTEIN E; NONSTRUCTURAL
 DE PROTEINS NS1, NS2A, NS2B, NS4A AND NS4B; HELICASE (NS3); RNA-DIRECTED
 DE RNA POLYMERASE (EC 2.7.7.48) (NS5)).
 OS TICK-BORNE ENCEPHALITIS VIRUS (STRAIN SOFJIN) (TBEV).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; FLAVIVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90101381
 RA PLETNEV A.G., YAMSHCHIKOV V.F., BLINOV V.M.;
 RL VIROLOGY 174:250-263(1990).
 RN [2]
 RP SEQUENCE OF 1-1190 FROM N.A.
 RM 88319988
 RA YAMSHCHIKOV V.F., PLETNEV A.G.;
 RL NUCLEIC ACIDS RES. 16:7750-7750(1988).
 RN [3]
 RP SEQUENCE OF 1-683 AND 758-1002 FROM N.A.
 RM 86220766
 RA PLETNEV A.G., YAMSHCHIKOV V.F., BLINOV V.M.;
 RL FEBS LETT. 200:317-321(1986).
 CC -!- FUNCTION: THE SMALL PROTEINS NS2A, NS2B, NS4A AND NS4B ARE
 CC HYDROPHOBIC, SUGGESTING A POSSIBLE MEMBRANE-RELATED FUNCTION.
 CC NS3 AND NS5 MAY PLAY A ROLE IN THE VIRAL RNA REPLICATION.
 CC -!- SUBUNIT: THE VIRION OF THIS VIRUS IS A NUCLEOCAPSID COVERED BY A
 CC LIPOPROTEIN ENVELOPE. THE ENVELOPE CONSISTS OF TWO PROTEINS:
 CC PROTEIN M AND GLYCOPROTEIN E. THE NUCLEOCAPSID IS A COMPLEX OF
 CC PROTEIN C AND MRNA.
 CC -!- THE NONSTRUCTURAL PROTEINS NS1 PRESENTS TWO ALTERNATIVE CLEAVAGE
 CC SITES FOR ITS C-TERMINUS, WHICH MAY DEFINE A SOLUBLE OR A
 CC MEMBRANE-BOUND FORM OF NS1.
 DR EMBL; X07755; TBEV1.
 DR EMBL; X03870; TOTBEV1.
 DR EMBL; X03871; TOTBEV2.
 DR PIR; A33776; GNWVTB.
 DR PIR; A24055; GNWVTE.
 DR PIR; B24055; MNWVTE.
 KW POLYPROTEIN; GLYCOPROTEIN; RNA-DIRECTED RNA POLYMERASE; CORE PROTEIN;
 KW COAT PROTEIN; ENVELOPE PROTEIN; HELICASE; ATP-BINDING; TRANSMEMBRANE;
 KW NONSTRUCTURAL PROTEIN.
 FT INIT_MET 1 1 REMOVED FROM CAPSID PROTEIN C BY THE
 FT CELLULAR AMINOPEPTIDASE.
 FT CHAIN 1 1 CAPSID PROTEIN C

FT	CHAIN	206	280	ENVELOPE GLYCOPROTEIN M.
FT	CHAIN	281	776	MAJOR ENVELOPE PROTEIN E.
FT	CHAIN	777	21128	NONSTRUCTURAL PROTEIN NS1 (OR 1190).
FT	CHAIN	21129	1358	NONSTRUCTURAL PROTEIN NS2A (OR 1191).
FT	CHAIN	1359	1489	NONSTRUCTURAL PROTEIN NS2B.
FT	CHAIN	1490	2110	HELICASE (NS3).
FT	CHAIN	2111	2259	NONSTRUCTURAL PROTEIN NS4A.
FT	CHAIN	2260	2510	NONSTRUCTURAL PROTEIN NS4B.
FT	CHAIN	2511	3412	RNA-DIRECTED RNA POLYMERASE (NS5).
FT	TRANSMEM	101	112	HYDROPHOBIC SIGNAL SEQUENCE (POTENTIAL).
FT	TRANSMEM	247	259	POTENTIAL.
FT	TRANSMEM	266	280	POTENTIAL.
FT	TRANSMEM	738	751	POTENTIAL.
FT	CARBOHYD	144	144	POTENTIAL.
FT	CARBOHYD	434	434	
FT	CARBOHYD	861	861	POTENTIAL.
FT	CARBOHYD	983	983	POTENTIAL.
FT	CARBOHYD	999	999	POTENTIAL.
FT	CARBOHYD	1228	1228	POTENTIAL.
FT	CARBOHYD	2447	2447	POTENTIAL.
FT	CARBOHYD	2466	2466	POTENTIAL.
FT	CONFLICT	381	381	W -> S (IN REF. 3).
FT	CONFLICT	850	850	E -> D (IN REF. 3).
SQ	SEQUENCE	3412 AA; 377976 MW; 2.068228E+07 CN;		
CC	-!- Retrieved by alexk on Thu 25 Feb 93 10:21:17-PST using FastDB			

Initial Score = 6 Optimized Score = 6 Significance = 3.09
Residue Identity = 100% Matches = 6 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

RRVTTASAAQRRGRVGRQEGRTDEYIYSGQDDDDSGLVQWKEAQILLDNITTLRGPVATFYGPQDKMPEV
1940 1950 1960 1970 1980 1990 2000 2010

X X
WHVAAN

AGHFRLTEEKRKHFRLHLLTHCDFTPWLAWHVAANVSSVTSRNWTWEGPEENTVDEANGDLVTFERSPNGAERT
2020 2030 2040 X 2050 2060 2070 2080

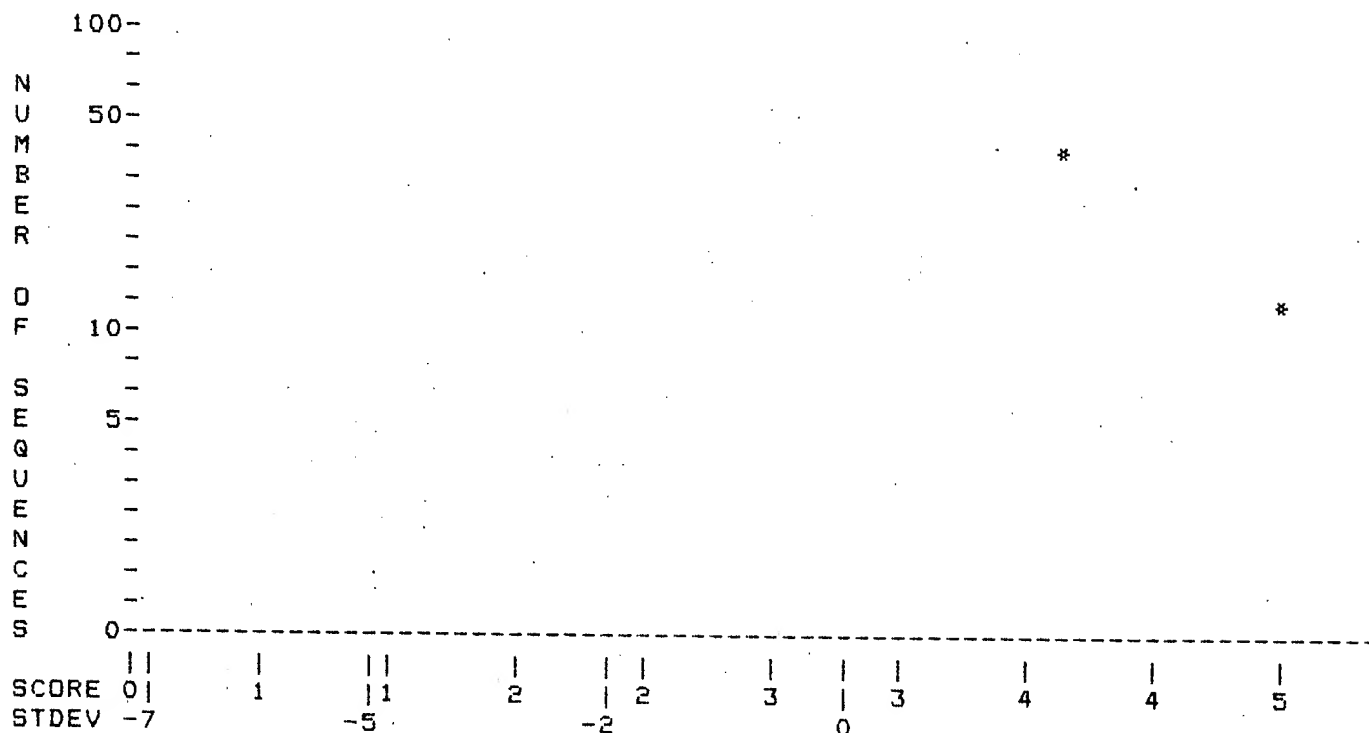
LRPVWRDARMFREGRDIREFVAYASGRRSFGDVLSGMSGVPELLRHRCVSAMDVFYTLMHHEE
2090 2100 2110 2120 2130 2140

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-13-align.res made by alexk on Thu 25 Feb 93 10:47:27-PST.

Query sequence being compared: CELSA-13 (1-5)
Number of sequences searched: 50
Number of scores above cutoff: 50

Results of the initial comparison of CELSA-13 (1-5) with:
File : celsa-13.pep



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	5	0.42

Times:	CPU	Total Elapsed
	00:00:00.02	00:00:01.00

Number of residues:	15270
Number of sequences searched:	50
Number of scores above cutoff:	50

Significance is calculated based on initial score.

11 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. A37113	*Ryanodine receptor, cardiac	4969	5	5	2.39	0
2. BVFFSL	sol protein, large splice for	1597	5	5	2.39	0
3. SDL_DROME	SMALL OPTIC LOBES PROTEIN.	1597	5	5	2.39	0
4. A31854	Cytochrome P450 51 lanosterol	528	5	5	2.39	0
5. CP51_CANTR	CYTOCHROME P450 L1 (P450-L1A1	528	5	5	2.39	0
6. STP1_ARATH	GLUCOSE TRANSPORTER (SUGAR CA	522	5	5	2.39	0
7. S14627	*Glucose transport protein -	522	5	5	2.39	0
8. S12042	*Sugar transport protein STP1	522	5	5	2.39	0
9. B35901	*Calcium channel alpha-1 chai	164	5	5	2.39	0
10. J00700	Hypothetical 11K protein (mmo	103	5	5	2.39	0
11. YMMO_METCA	HYPOTHETICAL 11.9 KD PROTEIN	103	5	5	2.39	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
12. C24735	Glutathione transferase, 2-2	9	4	4	0.00	0
13. B24735	Glutathione transferase, 1-2	18	4	4	0.00	0
14. D24735	Glutathione transferase, 2-2	19	4	4	0.00	0
15. A24735	Glutathione transferase, 1-1	26	4	4	0.00	0
16. S21278	*Glutathione transferase chai	28	4	4	0.00	0
17. S09585	*Glutathione transferase - Ra	31	4	4	0.00	0
18. S03358	*Glutathione transferase - Ra	31	4	4	0.00	0
19. J00099	Hypothetical 7K protein - Pap	68	4	4	0.00	0
20. V07K_PMV	7 KD PROTEIN (ORF 4).	68	4	4	0.00	0

1. CELSA-13 (1-5)

A37113 *Ryanodine receptor, cardiac muscle - Rabbit

ENTRY A37113 #Type Protein
 TITLE *Ryanodine receptor, cardiac muscle - Rabbit
 DATE 12-Feb-1991 #Sequence 12-Feb-1991 #Text 12-Feb-1991
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT *This entry is not verified.
 SOURCE Oryctolagus cuniculus #Common-name domestic rabbit
 REFERENCE

#Authors Otsu K., Willard H.F., Khanna V.K., Zorzato F.,
 Green N.M., MacLennan D.H.
 #Journal J. Biol. Chem. (1990) 265:13472-13483
 #Title Molecular cloning of cDNA encoding the Ca(2+)
 release channel (ryanodine receptor) of rabbit
 cardiac muscle sarcoplasmic reticulum.

#Reference-number A37113

#Accession A37113

SUMMARY #Molecular-weight 565069 #Length 4969 #Checksum 5421

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:50-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

DLTSSDTFKEYDPDGKGIISKRDFHKAMESHKHYTQSETEFLLSCAETDENETLDYEEFVKRFHEPAKDIGF
 4030 4040 4050 4060 4070 4080 4090

V L N Y F
I I I I I

NVAVLLTNLSEHMPNDTRLQTFLELAESVLNYFQPF	LGRIEIMGSAKRIERVYFEISESSRTQWEKPQVKES
4100	4160

KRQIFDVVNEGGEKEKMELVNFCEDTIFEMQLAAQISESDLNERSANKEESEKERPEEQ
4170

2. CELSA-13 (1-5)

BVFFSL sol protein, large splice form - Fruit fly

ENTRY BVFFSL #Type Protein
 TITLE sol protein, large splice form - Fruit fly
 (Drosophila melanogaster)
 DATE 30-Jun-1992 #Sequence 30-Jun-1992 #Text 30-Jun-1992
 PLACEMENT 1221.0 1.0 1.0 1.0 1.0
 SOURCE Drosophila melanogaster
 ACCESSION A41146
 REFERENCE

#Authors Delaney S.J., Hayward D.C., Barleben F., Fischbach
 K.F., Miklos G.L.G.
 #Journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:7214-7218
 #Title Molecular cloning and analysis of small optic lobes,
 a structural brain gene of Drosophila
 melanogaster.
 #Reference-number A41146
 #Accession A41146
 #Molecule-type mRNA
 #Residues 1-1597
 #Cross-reference GB:M64084

COMMENT The sol (small optic lobes) mutation eliminates
 certain classes of columnar neurons.

COMMENT An alternate splice form of 395 amino acids is
 observed, in which the first 393 are identical to
 the large sol protein.

GENETIC

#Segment-number 19F4
 #Name sol

SUPERFAMILY #Name sol protein

FEATURE

1017-1320	#Domain calpain catalytic domain homology <CAL>\
10-29	#Region zinc finger-like motif <ZN1>\
139-158	#Region zinc finger-like motif\
647-667	#Region zinc finger-like motif\
711-730	#Region zinc finger-like motif\
752-771	#Region zinc finger-like motif\
934-953	#Region zinc finger-like motif\
1082	#Active-site Cys (predicted)\
1248	#Active-site His (predicted)\
1268	#Active-site Asn (predicted)

SUMMARY #Molecular-weight 174713 #Length 1597 #Checksum 8253

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

Initial Score	=	5	Optimized Score	=	5	Significance	=	2.39
Residue Identity	=	100%	Matches	=	5	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

AQLSSRCVRF	LMGASCGG	GNMKVDEE	EYQQKGLR	PRHAYS	VL	DVKD	IQGH	RL	KL	RNPW	GHYS	WRGD	WSD
1220	1230	1240	1250	1260	1270	1280							

X X
VLNYF
I I I I I

 1290 1300 1310 X 1320 1330 1340 1350

AEFTLFQEGQRNSEKSRQSLDLCVVIFRTRSPAAPFGRVLEHRSKRQVRGFGVGHKMLER
 1360 1370 1380 1390 1400 1410

3. CELSA-13 (1-5)

SOL_DROME SMALL OPTIC LOBES PROTEIN.

ID SOL_DROME STANDARD; PRT; 1597 AA.
 AC P27398;
 DT 01-AUG-1992 (REL. 23, CREATED)
 DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE SMALL OPTIC LOBES PROTEIN.
 GN SOL.
 OS DROSOPHILA MELANOGASTER (FRUIT FLY).
 OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=BRAIN;
 RM 91334436
 RA DELANEY S.J., HAYWARD D.C., BARLEBEN F., FISCHBACH K.F.,
 RA MIKLOS G.L.G.;
 RL PROC. NATL. ACAD. SCI. U.S.A. 88:7214-7218(1991).
 CC -!- THE SOL (SMALL OPTIC LOBES) MUTATION ELIMINATES CERTAIN CLASSES OF
 CC COLUMNAR NEURONS.
 CC -!- ALTERNATIVE SPLICING: AN ALTERNATE SPLICE FORM OF 395 AMINO ACIDS
 CC IS OBSERVED, IN WHICH THE FIRST 393 ARE IDENTICAL TO THE LARGE SOL
 CC PROTEIN.
 CC -!- SIMILARITY: TO CALPAINS EUKARYOTIC THIOL PROTEASES.
 DR EMBL; M64084; DMSOL.
 DR PIR; A41146; BVFFSL.
 DR FLYBASE; 03464; RELEASE 9206.
 KW ALTERNATIVE SPLICING; ZINC-FINGER; HYDROLASE; THIOL PROTEASE.
 FT ZN_FING 10 29 C4-TYPE.
 FT ZN_FING 139 158 C4-TYPE.
 FT ZN_FING 647 667 C4-TYPE.
 FT DOMAIN 673 689 GLN-RICH.
 FT DOMAIN 690 697 POLY-HIS.
 FT ZN_FING 711 730 C4-TYPE.
 FT ZN_FING 752 771 C4-TYPE.
 FT ZN_FING 934 953 C4-TYPE.
 FT DOMAIN 1017 1320 CALPAIN CATALYTIC DOMAIN.
 FT ACT_SITE 1082 1082 BY SIMILARITY.
 FT ACT_SITE 1248 1248 BY SIMILARITY.
 FT ACT_SITE 1268 1268 BY SIMILARITY.
 SQ SEQUENCE 1597 AA; 174714 MW; 1.268343E+07 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

AQLSSRCVRFLMGASCGGGMKVDEEEYQQKGLRPRHAYSVLDVKDIQGHRLCLKLRNPWGHYSWRGDWSDD
 1220 1230 1240 1250 1260 1270 1280

X X
 VLNYF

||||

SSLWTDDLRLDALMPHGASEGVFWISFEDVLNYFDCIDICKVRSQWNEVRLQGTLOPLCSISCVLLTVLEPTE
 1290 1300 1310 X 1320 1330 1340 1350

AEFTLFQEGQRNSEKSRQSLDLCVVIFRTRSPAAPFGRVLEHRSKRQVRGFGVGHKMLER

4. CELSA-13 (1-5)

A31854 Cytochrome P450 51 lanosterol 14alpha-demethylase

ENTRY A31854 #Type Protein
 TITLE Cytochrome P450 51 lanosterol 14alpha-demethylase -
 Imperfect fungus (Candida tropicalis)
 ALTERNATE-NAME cytochrome P450 14DM
 DATE 07-Jun-1990 #Sequence 07-Jun-1990 #Text 10-Aug-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 SOURCE Candida tropicalis
 ACCESSION A31854
 REFERENCE
 #Authors Chen C., Kalb V.F., Turi T.G., Loper J.C.
 #Journal DNA (1988) 7:617-626
 #Title Primary structure of the cytochrome P450 lanosterol
 14alpha-demethylase gene from Candida tropicalis.
 #Reference-number A31854
 #Accession A31854
 #Molecule-type DNA
 #Residues 1-528 <CHE>
 SUPERFAMILY #Name cytochrome P450
 KEYWORDS heme\ monooxygenase\ oxidoreductase
 FEATURE
 470 #Binding-site heme iron (Cys) (axial
 ligand)
 SUMMARY #Molecular-weight 60927 #Length 528 #Checksum 4431
 SEQUENCE
 COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

MQPYEFFEKCRLKYGDVFSFMLLGKVM TVYLGPKGHEFIYNAKLSDVSAEEAYTHLTTPVFGKGVIYDCPNS
 70 80 90 100 110 120 130

X X

VLNYF

IIII

RLMEQKKFAKFALTTDSFKTYVPKIREEV LN YFVNDVSFKTKERDHGVASVMKTQPEITIFTASRCLFGDEM
 140 150 160 X 170 180 190 200

RKSFDRSFAQLYADLDKGFTPINFVFPNLPLPHYWRRDAAQRKISAHYMKIKRRRESGDI
 210 220 230 240 250 260 270

5. CELSA-13 (1-5)

CP51_CANTR CYTOCHROME P450 L1 (P450-L1A1) (LANOSTEROL 14-ALPH

ID CP51_CANTR STANDARD; PRT; 528 AA.
 AC P14263;
 DT 01-JAN-1990 (REL. 13, CREATED)
 DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
 DT 01-NOV-1990 (REL. 16, LAST ANNOTATION UPDATE)
 DE CYTOCHROME P450 L1 (P450-L1A1) (LANOSTEROL 14-ALPHA DEMETHYLASE)
 DE (EC 1.14.14.1).
 GN CYP51 OR 14DM.
 OS CANDIDA TROPICALIS (YEAST).
 OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 89152749
 RA CHEN C., KALB V.F., TURI T.G., LOPEL J.C.;

RP SEQUENCE OF 434-528 FROM N.A.
 RM 87298576
 RA CHEN C., TURI T.G., SANGLARD D., LOPER J.C.;
 RL BIOCHEM. BIOPHYS. RES. COMMUN. 146:1311-1317(1987).
 CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
 CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
 CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
 CC -!- CATALYTIC ACTIVITY: 14-ALPHA-DEMETHYLATION OF LANOSTEROL.
 DR EMBL; M23673; M23673.
 DR EMBL; M17595; CT14DM.
 DR PIR; A31854; A31854.
 DR PIR; A26828; A26828.
 DR PROSITE; PS00086; CYTOCHROME_P450.
 KW ELECTRON TRANSPORT; OXIDOREDUCTASE; MONOOXYGENASE; MEMBRANE;
 KW HEME.
 FT BINDING 470 470 HEME.
 SQ SEQUENCE 528 AA; 60928 MW; 1513726 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

MQPYEFFEKCRLKYGDVFSFMLLGKVM TVYLGPKGHEFIYNAKLSDVSAEEAYTHLTTPVFGKGVIYDCPNS
 70 80 90 100 110 120 130

X X
 VLNYF

||||

RLMEQKKFAKFALTTDSFKTYVPKIREEVNLNYFVNDVSFKTKERDHGVASVMKTQFEITIFTASRCLFGDEM
 140 150 160 X 170 180 190 200

RKSFDRSFAQLYADLDKGFTPINFVFPNLPLPHYWRRDAAQRKISAHYMKIKRRRESGDI
 210 220 230 240 250 260 270

6. CELSA-13 (1-5)

STP1_ARATH GLUCOSE TRANSPORTER (SUGAR CARRIER).

ID STP1_ARATH STANDARD; PRT; 522 AA.
 AC P23586;
 DT 01-NOV-1991 (REL. 20, CREATED)
 DT 01-NOV-1991 (REL. 20, LAST SEQUENCE UPDATE)
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
 DE GLUCOSE TRANSPORTER (SUGAR CARRIER).
 GN STP1.
 OS ARABIDOPSIS THALIANA (MOUSE-EAR CRESS).
 OC EUKARYOTA; PLANTA; EMBRYOPHYTA; ANGIOSPERMAE; DICOTYLEDONEAE;
 OC CAPPARALES; CRUCIFERAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. LANDSBERG ERECTA;
 RM 91005995
 RA SAUER N., FRIEDLAENDER K., GRAEML-WICKE U.;
 RL EMBD J. 9:3045-3050(1990).
 CC -!- FUNCTION: ACTIVE UPTAKE OF HEXOSES. PROBABLE GLUCOSE/HYDROGEN
 CC SYMPORT.
 CC -!- SIMILARITY: TO OTHER EUKARYOTIC AND PROKARYOTIC SUGAR SYMPORTERS.
 DR EMBL; X55350; ATSTP1.
 DR PIR; S12042; S12042.
 DR PIR; S14627; S14627.
 DR PROSITE; PS00216; SUGAR_TRANSPORT_1.
 DR PROSITE; PS00217; SUGAR_TRANSPORT_2.
 KW DUPLICATION; TRANSMEMBRANE; SUGAR TRANSPORT; SYMPORT.
 FT

```

FT TRANSMEM 113 133 POTENTIAL.
FT TRANSMEM 141 161 POTENTIAL.
FT TRANSMEM 168 188 POTENTIAL.
FT TRANSMEM 200 220 POTENTIAL.
FT TRANSMEM 285 305 POTENTIAL.
FT TRANSMEM 321 341 POTENTIAL.
FT TRANSMEM 349 369 POTENTIAL.
FT TRANSMEM 385 405 POTENTIAL.
FT TRANSMEM 424 444 POTENTIAL.
FT TRANSMEM 453 473 POTENTIAL.
SQ SEQUENCE 522 AA; 57596 MW; 1520081 CN;
CC -!- Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

```

```

Initial Score      =      5   Optimized Score =      5   Significance = 2.39
Residue Identity   =    100%   Matches         =      5   Mismatches   =      0
Gaps               =      0   Conservative Substitutions =      0

```

```

SSLYLAALISLVASTVTRKFGRRLSMLFGGILFCAGALINGFAKHVWMLIVGRILLGFGIGFANQAVPLYL
 90      100      110      120      130      140      150

                        X   X
                        VLNYF
                        |||||
SEMAPYKYRGALNIGFQLSITIGILVAEVLNYFFAKIKGGGWRLSLGGAVVPALIITIGSLVLPDTPNSMI
160      170      180      190      200      210      220      230

ERGQHEEAKTKLRRIRGVDDVSQEFDDLVAASKESQSIHPWRNLLRRKYRPHLTMAVMIP
 240      250      260      270      280      290

```

7. CELSA-13 (1-5)

S14627 *Glucose transport protein - Arabidopsis thaliana

```

ENTRY      S14627      #Type Protein
TITLE      *Glucose transport protein - Arabidopsis thaliana
DATE       28-Aug-1992 #Sequence 28-Aug-1992 #Text 28-Aug-1992
PLACEMENT  0.0      0.0      0.0      0.0      0.0
COMMENT    *This entry is not verified.
SOURCE     Arabidopsis thaliana #Common-name mouse-ear cress
REFERENCE

#Authors   Sauer N., Friedl K., Wicke U.
#Citation  submitted to the EMBL Data Library, October 1990
#Title     Primary structure, genomic organization &
           heterologous expression of a glucose transporter
           from Arabidopsis thaliana.
#Reference-number S14627
#Accession  S14627
#Cross-reference EMBL:X55350

```

SUMMARY *Molecular-weight 57596 #Length 522 #Checksum 5679

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

```

Initial Score      =      5   Optimized Score =      5   Significance = 2.39
Residue Identity   =    100%   Matches         =      5   Mismatches   =      0
Gaps               =      0   Conservative Substitutions =      0

```

```

SSLYLAALISLVASTVTRKFGRRLSMLFGGILFCAGALINGFAKHVWMLIVGRILLGFGIGFANQAVPLYL
 90      100      110      120      130      140      150

                        X   X
                        VLNYF
                        |||||
SEMAPYKYRGALNIGFQLSITIGILVAEVLNYFFAKIKGGGWRLSLGGAVVPALIITIGSLVLPDTPNSMI
160      170      180      190      200      210      220      230

```

8. CELSA-13 (1-5)

S12042

*Sugar transport protein STP1 - Arabidopsis thaliana

ENTRY S12042 #Type Protein
 TITLE *Sugar transport protein STP1 - Arabidopsis thaliana
 DATE 26-Sep-1992 #Sequence 26-Sep-1992 #Text 26-Sep-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT *This entry is not verified.
 SOURCE Arabidopsis thaliana #Common-name mouse-ear cress
 REFERENCE
 #Authors Sauer N., Friedlaender K., Graenl-Wicke U.
 #Journal EMBO J. (1990) 9:3045-3050
 #Title Primary structure, genomic organization and heterologous expression of a glucose transporter from Arabidopsis thaliana.
 #Reference-number S12042
 #Accession S12042
 #Cross-reference EMBL:X55350
 SUMMARY #Molecular-weight 57596 #Length 522 #Checksum 5679
 SEQUENCE
 COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:51-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

SSLYLAALISLVASTVTRKFGRRLSMLFGGILFCAGALINGFAKHVWMLIVGRILLGFGIGFANQAVPLYL
 90 100 110 120 130 140 150

X X
 VLNYF
 |||||

SEMAPYKYRGALNIGFQLSITIGILVAEVLNYFFAKIKGGWGWRLSLGGAVVPALIITIGSLVLPDTPNSMI
 160 170 180 190 200 210 220 230

ERGQHEEAKTKLRRIRGVDDVDSQEFDDLVAASKESQSIHPWRNLLRRKYRPHLTMVAVMIP
 240 250 260 270 280 290

9. CELSA-13 (1-5)

B35901

*Calcium channel alpha-1 chain, dihydropyridine

ENTRY B35901 #Type Protein (fragment)
 TITLE *Calcium channel alpha-1 chain, dihydropyridine sensitive, homolog B, brain - Rat (fragment)
 DATE 18-Apr-1991 #Sequence 18-Apr-1991 #Text 18-Apr-1991
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT *This entry is not verified.
 SOURCE Rattus norvegicus #Common-name Norway rat
 REFERENCE
 #Authors Snutch T.P., Leonard J.P., Gilbert M.M., Lester H.A., Davidson N.
 #Journal Proc. Natl. Acad. Sci. U.S.A. (1990) 87:3391-3395
 #Title Rat brain expresses a heterogeneous family of calcium channels.
 #Reference-number B35901
 #Accession B35901
 SUMMARY #Length 164 #Checksum 2550
 SEQUENCE
 COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:52-PST using FastDB

Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

      X   X
      VLNYF
      |||||
FSLECILKIIAFQVLNYFRDAWVDFVTVLGSITDILVTEIANNFINLSFLRLFRAARLIKLLRQGYTIRI
      10   X   20           30           40           50           60           70

LLWTFVQSFKALPYVCLLIAMLFFIYAIIGMQVFGNIALDDGTSIN
      80           90           100          110

```

10. CELSA-13 (1-5)

J00700 Hypothetical 11K protein (mmoC 5' region) -

ENTRY J00700 #Type Protein
 TITLE Hypothetical 11K protein (mmoC 5' region) -
 Methylococcus capsulatus
 ALTERNATE-NAME hypothetical protein Y
 DATE 31-Mar-1992 #Sequence 31-Mar-1992 #Text 31-Mar-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 SOURCE Methylococcus capsulatus
 ACCESSION J00700
 REFERENCE

#Authors Stainthorpe A.C., Lees V., Salmond G.P.C., Dalton
 H., Murrell J.C.

#Journal Gene (1990) 91:27-34

#Title The methane monooxygenase gene cluster of
 Methylococcus capsulatus (Bath).

#Reference-number J00700

#Accession J00700

#Molecule-type DNA

#Residues 1-103 <STA>

SUMMARY #Molecular-weight 11942 #Length 103 #Checksum 8171

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:39:52-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

MVESAFQPFSGDADEWFEEPRPQAGFFPSADWHLKRDETYAAYAKDLDFMWRWVIVREERIVQEGCSISLE
      10           20           30           40           50           60           70

      X   X
      VLNYF
      |||||
SSIRAVTHVLNYFGMTEQRAPAEDRTGGVQH
      80   X   90           100

```

11. CELSA-13 (1-5)

YMMO_METCA HYPOTHETICAL 11.9 KD PROTEIN IN MMOZ-MMOC INTERGEN

ID YMMO_METCA STANDARD; PRT; 103 AA.
 AC P22867;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-NOV-1991 (REL. 20, LAST ANNOTATION UPDATE)
 DE HYPOTHETICAL 11.9 KD PROTEIN IN MMOZ-MMOC INTERGENIC REGION (ORFY).
 OS METHYLOCOCCUS CAPSULATUS.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 OC METHYLOCOCCACEAE.
 RN [1]
 SS

RM 90382694
 RA STAINTHORPE A.C., LEES V., SALMOND G.P.C., DALTON H., MURRELL J.C.;
 RL GENE 91:27-34(1990).
 DR EMBL; M58498; MCMMDA.
 DR PIR; JQ0700; JQ0700.
 KW HYPOTHETICAL PROTEIN.
 SQ SEQUENCE 103 AA; 11942 MW; 52861 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:39:52-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.39
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

MVESAFQPFSGDADEWFEEPRPQAGFFPSADWHLLKRDETYAAYAKDLDFMWRWVIVREERIVQEGCSISLE
 10 20 30 40 50 60 70

X X

VLNYF

IIII

SSIRAVTHVLNYFGMTEQRAPAEDRTGGVQH

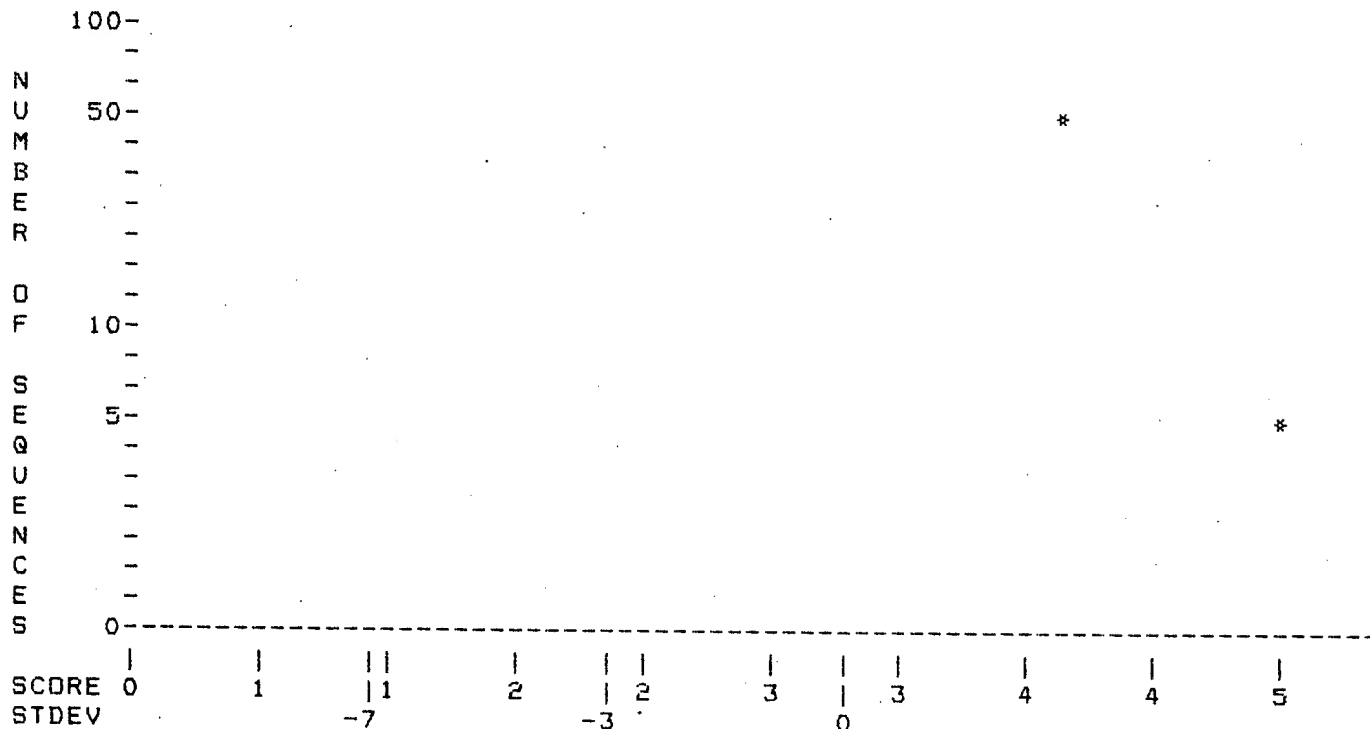
80 X 90 100

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-3-align.res made by alexk on Thu 25 Feb 93 10:48:08-PST.

Query sequence being compared: CELSA-3 (1-5)
Number of sequences searched: 50
Number of scores above cutoff: 50

Results of the initial comparison of CELSA-3 (1-5) with:
File : celsa-3.pep



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	5	0.30
Times:	CPU	Total Elapsed	
	00:00:00.01	00:00:00.00	
Number of residues:	8896		
Number of sequences searched:	50		
Number of scores above cutoff:	50		

5 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Opt.		Sig.	Frame
			Score	Score		
1. P82507	Pullulanase protein.	1096	5	5	3.30	0
2. A26879	alpha-Dextrin endo-1,6-alpha-	1096	5	5	3.30	0
3. PULA_KLEAE	PULLULANASE (EC 3.2.1.41) (AL	1096	5	5	3.30	0
4. S12888	*DNA-binding protein II - The	95	5	5	3.30	0
5. DBH_THETH	DNA-BINDING PROTEIN II.	95	5	5	3.30	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Opt.		Sig.	Frame
			Score	Score		
6. S10561	*Photosystem II 21K protein -	66	4	4	0.00	0
7. CCPS55	Cytochrome c551 - Pseudomonas	82	4	4	0.00	0
8. C551_PSEST	CYTOCHROME C551.	82	4	4	0.00	0
9. CCPB6	Cytochrome c6 - Plectonema bo	85	4	4	0.00	0
10. CYC6_PLEBO	CYTOCHROME C6 (SOLUBLE CYTOCH	85	4	4	0.00	0
11. CCAI6	Cytochrome c6 - Anabaena vari	86	4	4	0.00	0
12. CYC6_ANAVA	CYTOCHROME C6 (SOLUBLE CYTOCH	86	4	4	0.00	0
13. DNBS2F	DNA-binding protein II - Baci	90	4	4	0.00	0
14. DBH_BACST	DNA-BINDING PROTEIN II (HB) (90	4	4	0.00	0
15. S00015	DNA-binding protein HB - Baci	92	4	4	0.00	0
16. B39409	*DNA-binding protein HB - Bac	92	4	4	0.00	0
17. DBH_BACSU	DNA-BINDING PROTEIN II (HB) (92	4	4	0.00	0
18. R05446	CAT-GLP-1 hybrid protein.	104	4	4	0.00	0
19. QQECRP	Hypothetical protein A-105 -	105	4	4	0.00	0
20. S06985	Hypothetical protein (nifH1 3	105	4	4	0.00	0

1. CELSA-3 (1-5)

P82507 Pullulanase protein.

ID P82507 standard; protein; 1096 AA.
AC P82507;
DT 01-NOV-1990 (first entry)
DE Pullulanase protein.
KW Pullulanase; starch; alcohol prodn.
OS Klebsiella aerogenes.
PN J63245676-A.
PD 12-OCT-1988.
PF 31-MAR-1987; 078355.
PR 31-MAR-1987; JP-078355.
PA (ELED) Denki Kagaku Kogyo KK; (SUNR) Suntory Ltd.
DR WPI; 88-333488/47.
DR N-PSDB; N81341.
PT Gene encoding pullulanase - derived from recombinant plasmid pMP1 contg.
PT gene from Klebsiella genus.
PS Disclosure; p; Japanese.
CC The pullulanase protein cleaves alpha-1,6-glucoside bonds of starch and
CC is effective in decomposition of branched starch. It is used in the
CC prodn. of maltose and glucose from starch, and of alcohol from starch
CC via glucose. Amino acid residues 1-19 can be deleted.
SQ Sequence 1096 AA;
SQ 108 A; 63 R; 42 N; 93 D; 0 B; 10 C; 51 Q; 42 E; 0 Z; 89 G; 22 H;
SQ 38 I; 86 L; 42 K; 21 M; 31 F; 52 P; 96 S; 70 T; 18 W; 41 Y; 81 V;
CC Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 3.30

```

SFSDFTDRTVSVIAGNSAVYDSRADAFRAAFGVALADAHWVDKTTLLWPGGENKPIVRLYYSHSSKVAADSN
160      170      180      190      200      210      220      230
          X      X
          YPAFK
          |||||
GEFSDKYVKLTPTTVNQVSMRFPHLASYPAFKLPDDVNVDELLOGDDGGIAESDGILSLSHPGADRRRAGR
240      250      260      X      270      280      290      300
          YLCRRAEALSYGAQLTDSGVTFRVWAPTAQQVELVIYSADKKVIA SHPMTRDSASGAWSWQ
310      320      330      340      350      360

```

2. CELSA-3 (1-5)

A26879 alpha-Dextrin endo-1,6-alpha-glucosidase precursor

```

ENTRY      A26879      #Type Protein
TITLE      alpha-Dextrin endo-1,6-alpha-glucosidase precursor -
            Klebsiella pneumoniae #EC-number 3.2.1.41
ALTERNATE-NAME pullulanase
DATE       30-Jun-1988 #Sequence 30-Jun-1988 #Text 31-Sep-1992
PLACEMENT  0.0      0.0      0.0      0.0      0.0
SOURCE     Klebsiella pneumoniae
ACCESSION  A26879
REFERENCE  (K. aerogenes, strain W70)
            #Authors   Katsuragi N., Takizawa N., Murooka Y.
            #Journal   J. Bacteriol. (1987) 169:2301-2306
            #Title     Entire nucleotide sequence of the pullulanase gene
                     of Klebsiella aerogenes W70.
            #Reference-number A26879
            #Accession  A26879
            #Molecule-type DNA
            #Residues   1-1096 <KAT>

```

GENETIC

```

#Name      pula
KEYWORDS   glycosidase\ hydrolase
FEATURE

```

```

1-19      #Domain signal sequence <SIG>\
20-1096   #Protein alpha-dextrin endo-1,
          6-alpha-glucosidase <MAT>

```

SUMMARY #Molecular-weight 119335 #Length 1096 #Checksum 1390

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

```

Initial Score = 5 Optimized Score = 5 Significance = 3.30
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

```

```

SFSDFTDRTVSVIAGNSAVYDSRADAFRAAFGVALADAHWVDKTTLLWPGGENKPIVRLYYSHSSKVAADSN
160      170      180      190      200      210      220      230
          X      X
          YPAFK
          |||||
GEFSDKYVKLTPTTVNQVSMRFPHLASYPAFKLPDDVNVDELLOGDDGGIAESDGILSLSHPGADRRRAGR
240      250      260      X      270      280      290      300
          YLCRRAEALSYGAQLTDSGVTFRVWAPTAQQVELVIYSADKKVIA SHPMTRDSASGAWSWQ
310      320      330      340      350      360

```

3. CELSA-3 (1-5)

PULA_KLEAE PULLULANASE (EC 3.2.1.41) (ALPHA-DEXTRIN ENDO-1,6-

AC P07811;
 DT 01-AUG-1988 (REL. 08, CREATED)
 DT 01-AUG-1988 (REL. 08, LAST SEQUENCE UPDATE)
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
 DE PULLULANASE (EC 3.2.1.41) (ALPHA-DEXTRIN ENDO-1,6-ALPHA-GLUCOSIDASE)
 DE (PULLULAN 6-GLUCANOHYDROLASE).
 GN PULA.
 OS KLEBSIELLA AEROGENES.
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; FACULTATIVELY ANAEROBIC RODS;
 OC ENTEROBACTERIACEAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=W70;
 RM 87194626
 RA KATSURAGI N., TAKIZAWA N., MUROOKA Y.;
 RL J. BACTERIOL. 169:2301-2306(1987).
 CC -!- CATALYTIC ACTIVITY: STARCH-DEBRANCHING ENZYME, HYDROLYZES
 CC (1-6)-ALPHA-GLUCOSIDIC LINKAGES IN PULLULAN AND STARCH TO
 CC FORM MALTOTRIOSE.
 CC -!- SUBUNIT: HOMOTRIMER.
 DR EMBL; M16187; KAPULA.
 DR PIR; A26879; A26879.
 DR PROSITE; PS00013; PROKAR_LIPOPROTEIN.
 KW HYDROLASE; GLYCOSIDASE; LIPOPROTEIN; SIGNAL.
 FT SIGNAL 1 19
 FT CHAIN 20 1096 PULLULANASE.
 FT LIPID 20 20 N-ACYL DIGLYCERIDE.
 SQ SEQUENCE 1096 AA; 119335 MW; 5852107 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 3.30
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

SFSDFTDRTVSVIAGNSAVYDSRADAFRAAFGVALADAHWVDKTTLLWPGGENKPIVRLYYSHSSKVAADSN
 160 170 180 190 200 210 220 230

 X X
 YPAFK
 I I I I
 GEFSDKYVKLTPTTVNQOVSMRFPHLASYPAFKLPDDVNVDELLOGDDGGIAESDGILSLSHPGADRRRAGR
 240 250 260 X 270 280 290 300

 YLCRRAEALSYGAQLTDSGVTFRVWAPTAQQVELVIYSADKKVIA SHPMTRDSASGAWSWQ
 310 320 330 340 350 360

4. CELSA-3 (1-5)

S12888 *DNA-binding protein II - Thermus aquaticus

ENTRY S12888 #Type Protein
 TITLE *DNA-binding protein II - Thermus aquaticus
 DATE 11-Apr-1992 #Sequence 11-Apr-1992 #Text 11-Apr-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT *This entry is not verified.
 SOURCE Thermus aquaticus
 REFERENCE
 #Authors Zierer R., Choli D.
 #Journal FEBS Lett. (1990) 273:59-62
 #Title The primary structure of DNA binding protein II from
 the extreme thermophilic bacterium Thermus
 thermophilus.
 #Reference-number S12888
 #Accession S12888

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 3.30
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

AAKKTVTKADLVDQVAQATGLKLLDVKAMVDALLAKVEEALANGSKVQLTGFGTFEVRKRKARTGVKPGTKE
 10 20 30 40 50 60 70

X X

YPAFK

|||||

KIKIPATQYPAFKPGKALKDKVK

80 X 90

5. CELSA-3 (1-5)

DBH_THETH DNA-BINDING PROTEIN II.

ID DBH_THETH STANDARD; PRT; 95 AA.
 AC P19436;
 DT 01-FEB-1991 (REL. 17, CREATED)
 DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
 DT 01-MAY-1991 (REL. 18, LAST ANNOTATION UPDATE)
 DE DNA-BINDING PROTEIN II.
 OS THERMUS AQUATICUS (SUBSP. THERMOPHILUS).
 OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
 OC UNCERTAIN.
 RN [1]
 RP SEQUENCE.
 RM 91032203
 RA ZIERER R., CHOLI D.;
 RL FEBS LETT. 273:59-62(1990).
 CC -!- FUNCTION: THIS PROTEIN BELONGS OT THE HISTONE LIKE FAMILY OF
 CC PROKARYOTIC DNA-BINDING PROTEINS WHICH ARE CAPABLE OF WRAPPING
 CC DNA AND TO STABILIZE DNA FROM DENATURATION UNDER EXTREME
 CC ENVIRONMENTAL CONDITIONS.
 CC -!- SUBUNIT: HOMODIMER.
 DR PIR; S12888; S12888.
 DR PROSITE; PS00045; HISTONE_LIKE.
 KW DNA-BINDING; DNA CONDENSATION.
 SQ SEQUENCE 95 AA; 10163 MW; 48066 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 3.30
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

AAKKTVTKADLVDQVAQATGLKLLDVKAMVDALLAKVEEALANGSKVQLTGFGTFEVRKRKARTGVKPGTKE
 10 20 30 40 50 60 70

X X

YPAFK

|||||

KIKIPATQYPAFKPGKALKDKVK

80 X 90

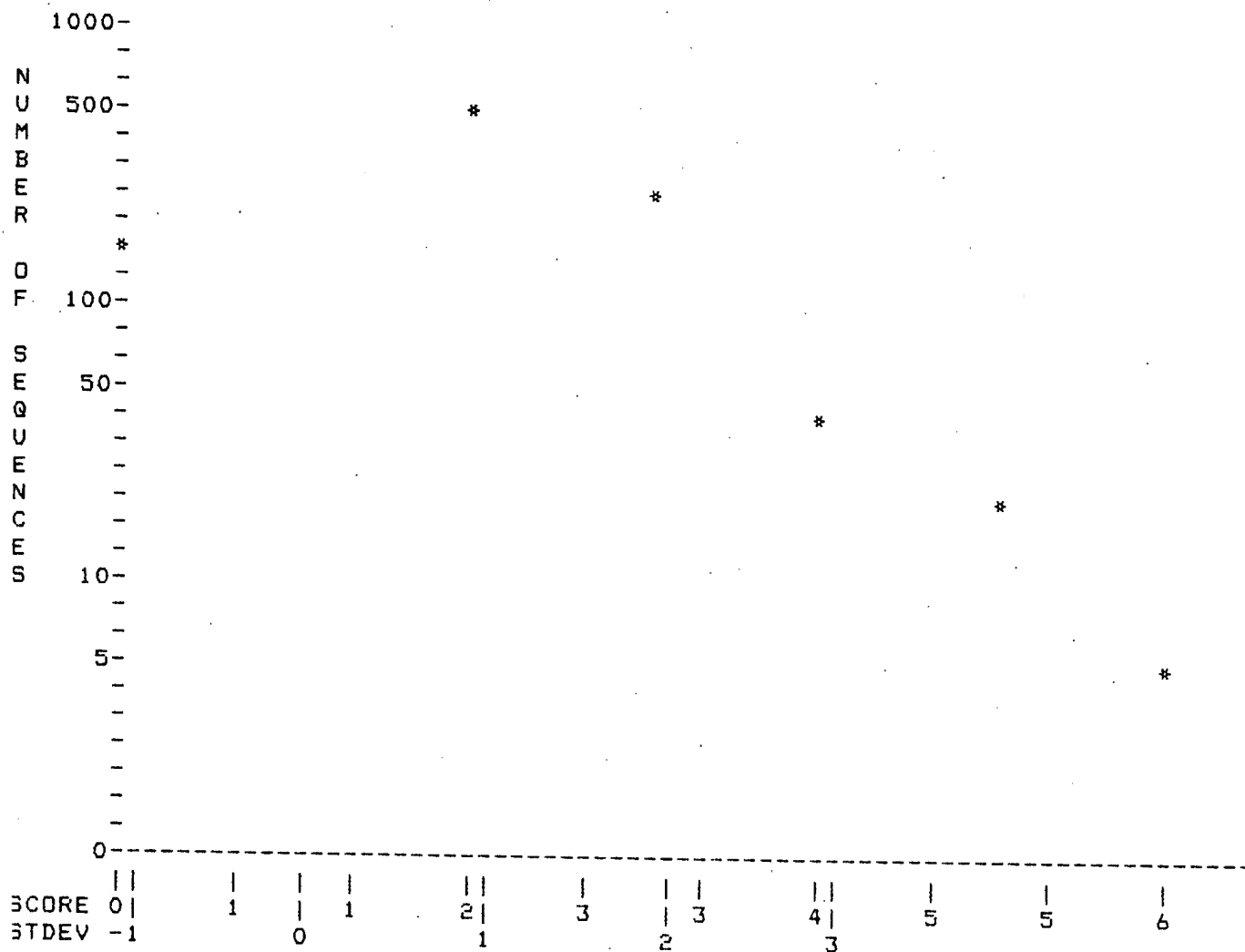
0| 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-align.res made by alexk on Thu 25 Feb 93 10:44:21-PST.

Query sequence being compared: CELSA-1 (1-6)
Number of sequences searched: 890 (interrupted)
Number of scores above cutoff: 757

Results of the initial comparison of CELSA-1 (1-6) with:
File : *.pep



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
---------	------	--------	--------------------

Times:

CPU
00:00:05.04

Total Elapsed
00:00:07.00

Number of residues:	442959
Number of sequences searched:	890
Number of scores above cutoff:	757

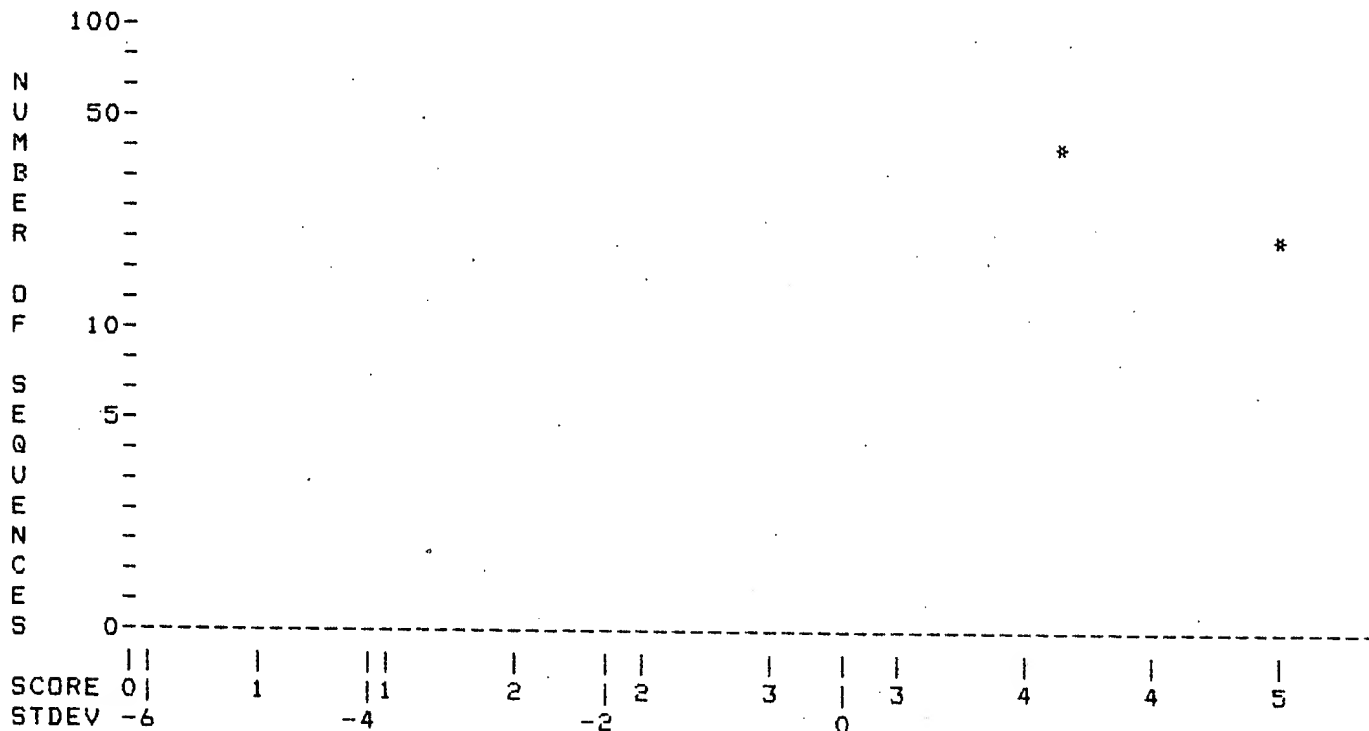
0| 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-4-align.res made by alexk on Thu 25 Feb 93 10:51:58-PST.

Query sequence being compared: CELSA-4 (1-5)
Number of sequences searched: 50
Number of scores above cutoff: 50

Results of the initial comparison of CELSA-4 (1-5) with:
File : celsa-4.pep



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	4	5	0.48

Times:	CPU	Total Elapsed
	00:00:00.01	00:00:00.00

Number of residues:	8503
Number of sequences searched:	50
Number of scores above cutoff:	50

Significance is calculated based on initial score.

17 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. S20590	*Sialidase - Actinomyces visc	913	5	5	2.09	0
2. A33827	*Regulatory protein ral2 - Ye	611	5	5	2.09	0
3. RAL2_SCHPO	RAL2 PROTEIN.	611	5	5	2.09	0
4. WMCVFM	Inclusion body matrix protein	512	5	5	2.09	0
5. IBMP_FMVD	INCLUSION BODY MATRIX PROTEIN	512	5	5	2.09	0
6. A30828	Steroid 17alpha-monooxygenase	507	5	5	2.09	0
7. S16719	*Steroid 17alpha-monooxygenas	507	5	5	2.09	0
8. CPT1_RAT	CYTOCHROME P450 XVIIIA1 (P450-	507	5	5	2.09	0
9. JH0594	Vasoactive intestinal polypep	459	5	5	2.09	0
10. S16562	*nolF protein - Rhizobium mel	367	5	5	2.09	0
11. NOLF_RHIME	NODULATION PROTEIN NOLF.	367	5	5	2.09	0
12. A27659	Cytochrome P450 17 - Rat (fra	237	5	5	2.09	0
13. A33980	Steroid 17alpha-monooxygenase	235	5	5	2.09	0
14. S18659	*Hypothetical protein - Mycop	148	5	5	2.09	0
15. R20793	CDR-grafted, humanised heavy	146	5	5	2.09	0
16. S06727	*Hypothetical protein 1 (mini	122	5	5	2.09	0
17. YM2_STRCD	MINI-CIRCLE HYPOTHETICAL 13.3	122	5	5	2.09	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
18. A35776	*RecQ protein - Escherichia c	5	4	4	0.00	0
19. S04171	aadA protein - Klebsiella pne	23	4	4	0.00	0
20. RP30_YEAST	RIBOSOMAL PROTEIN RP30 (FRAGM	25	4	4	0.00	0

1. CELSA-4 (1-5)

S20590 *Sialidase - Actinomyces viscosus #EC-number

ENTRY S20590 #Type Protein

TITLE *Sialidase - Actinomyces viscosus #EC-number
3.2.1.18

DATE 22-Jul-1992 #Sequence 22-Jul-1992 #Text 22-Jul-1992

PLACEMENT 0.0 0.0 0.0 0.0 0.0

COMMENT *This entry is not verified.

SOURCE Actinomyces viscosus

REFERENCE

#Authors Henningsen M., Roggentin P., Schauer R.

#Journal Biol. Chem. Hoppe-Seyler (1991) 372:1065-1072

#Title Cloning, sequencing and expression of the sialidase
gene from Actinomyces viscosus DSM 43798.

#Reference-number S20590

#Accession S20590

#Cross-reference EMBL:X62276

SUMMARY #Molecular-weight 96216 #Length 913 #Checksum 4303

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

PMGTCSSPTTSARRTTATAAATTPNPNHIVQRRSTDGGKTWSAPTYIHQGTETGKKVGYSDPSYVVDHQGTG
320 330 340 350 360 370 380 390

IFNFHVKSVDQGWGSGRGTDPENRGIIQAEVSTSTDNGWTWTHRTITADITKDKPWTFARFAASGGGIGIQH
 400 410 420 X 430 440 450 460

GPHAGRLVQOYTIRTAGGPVQAVSVYSDDHGKTWQAGTPIGTGMDENKVVELSDGSLMLNS
 470 480 490 500 510 520

2. CELSA-4 (1-5)

A33827 *Regulatory protein ral2 - Yeast

ENTRY A33827 #Type Protein
 TITLE *Regulatory protein ral2 - Yeast
 (Schizosaccharomyces pombe)
 DATE 19-Sep-1992 #Sequence 19-Sep-1992 #Text 19-Sep-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT *This entry is not verified.
 SOURCE Schizosaccharomyces pombe
 REFERENCE

#Authors Fukui Y., Miyake S., Satoh M., Yamamoto M.
 #Journal Mol. Cell. Biol. (1989) 9:5617-5622
 #Title Characterization of the Schizosaccharomyces pombe
 ral2 gene implicated in activation of the ras1
 gene product.

#Reference-number A33827

#Accession A33827

#Cross-reference GB:M30827

SUMMARY #Molecular-weight 69847 #Length 611 #Checksum 9734

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VVFGNKTRKLTQDYNLRQSNYDHIVFIELEGYGVYRKPMGRVTERSEQLGKLLNGISDMEILTIERMHIP
 380 390 400 410 420 430 440

X X
 QAEVS

CLSRMLYKRWPAFOKIMDRAVEKNQAEVQAEVSQLGPOQLTDLPFSSIHSTGSRALYMPYSFETCSAFLHYIY
 450 460 470 480 490 500 510 520

CGTLNGSYCTAKNLCNLLILCKGFEGLETFFAYIVHLLHGVLNRRNVKLIYETAALTGAKG
 530 540 550 560 570 580

3. CELSA-4 (1-5)

RAL2_SCHPD RAL2 PROTEIN.

ID RAL2_SCHPD STANDARD; PRT; 611 AA.
 AC P15258;
 DT 01-APR-1990 (REL. 14, CREATED)
 DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)
 DT 01-APR-1990 (REL. 14, LAST ANNOTATION UPDATE)
 DE RAL2 PROTEIN.
 GN RAL2.
 OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).
 OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90066514
 RA FUKUI Y., MIYAKE S., SATOH M., YAMAMOTO M.;

CC -!- FUNCTION: IMPLICATED IN ACTIVATION OF RAS1 PROTEIN.
 DR EMBL; M30827; SPRAL2.
 DR PIR; A33827; A33827.
 SQ SEQUENCE 611 AA; 69847 MW; 1910377 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

VVFQNKTRKLTQDYNLRQSNYDHIVFIELEGYGVYRKPMGRVTERSEQLGKLLNGISDMEILTIERMHP
380      390      400      410      420      430      440

                X   X
                QAEVS
                |||||
CLSRMLYKRWPAFQKIMDRAVEKNQEAFAEVSQLGPOLTDLPFSSIHSTGSRALYMPYSFETCSAFLHYIY
450      460      470      480      490      500      510      520

CGTLNGSYCTAKNLCNLLILCKGFEGLETFFAYIVHLLHGVLNRNNVKKLIYETAALTGAKG
530      540      550      560      570      580

```

4. CELSA-4 (1-5)

WMCVFM Inclusion body matrix protein - Figwort mosaic vir

ENTRY WMCVFM #Type Protein
 TITLE Inclusion body matrix protein - Figwort mosaic virus
 DATE 30-Sep-1991 #Sequence 30-Sep-1991 #Text 30-Jun-1992
 PLACEMENT 2377.0 1.0 1.0 1.0 1.0
 SOURCE figwort mosaic virus
 ACCESSION S01284
 REFERENCE

#Authors Richins R.D., Scholthof H.B., Shepherd R.J.
 #Journal Nucleic Acids Res. (1987) 15:8451-8466
 #Title Sequence of figwort mosaic virus DNA (caulimovirus group).

#Reference-number S01279

#Accession S01284

#Molecule-type DNA

#Residues 1-512 <RIC>

#Cross-reference EMBL:X06166

#Comment The translation of the nucleotide sequence is not given in this paper.

COMMENT This virus is a member of the group Caulimovirus.

SUPERFAMILY #Name caulimovirus inclusion body matrix protein

KEYWORDS matrix protein

SUMMARY #Molecular-weight 58207 #Length 512 #Checksum 1605

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

NLTPKSDKDKVKSSPVANGSGKDSTNPLNPVALGKSKMTILGQKQADEEEFKPDYLRAASNGQSWFAVYKGF
100      110      120      130      140      150      160

                X   X
                QAEVS
                |||||
NKEFFTWEIVADICKKRQKSKRFRSKEQAEVVISLYNKDIQDPVNFRLRPVKLVKEERAAQPLKFKAIAAEQ
170      180      190 X   X 200      210      220      230

TIQDFDEFROIWEKSRLSDLEDGVQEKFYTNDSASKSTYTFVENAEPYLVHTAFRAGLAKVI

```


5. CELSA-4 (1-5)

IBMP_FMVD INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN).

ID IBMP_FMVD STANDARD; PRT; 512 AA.
 AC P09524;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
 DE INCLUSION BODY MATRIX PROTEIN (VIROPLASMIN).
 GN VI.
 OS FIGWORT MOSAIC VIRUS (STRAIN DXS) (FMV).
 OC VIRIDAE; DS-DNA NONENVELOPED VIRUSES; CAULIMOVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 88040466
 RA RICHINS R.D., SCHOLTHOF H.B., SHEPHERD R.J.;
 RL NUCLEIC ACIDS RES. 15:8451-8466(1987).
 CC -!- FUNCTION: ENHANCES THE TRANSLATION OF DOWNSTREAM ORF'S ON
 CC POLYCISTRONIC MRNA'S DERIVED FROM FIGWORT MOSAIC VIRUS.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC INCLUSION BODIES.
 CC -!- THE INCLUSION BODIES ARE THE SITE OF VIRAL DNA SYNTHESIS, VIRION
 CC ASSEMBLY AND ACCUMULATION IN THE INFECTED CELL.
 CC -!- SIMILARITY: HIGH, WITH OTHER CAULIMOVIRUS VIROPLASMIN.
 DR EMBL; X06166; CAFMVXX.
 DR PIR; S01284; WMCVFM.
 KW TRANS-ACTING FACTOR; TRANSLATION REGULATION.
 SQ SEQUENCE 512 AA; 58207 MW; 1378914 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

NLTPKSDKDKVKSSPVANGSGKDSTNPLNPVALGKSKMTILGQKQADEEEFKPDYLRAASNGQSWFAVYKGP
 100 110 120 130 140 150 160
 X X
 QAEVS
 I I I I I
 NKEFFTEWEIVADICKKRQKSKRFRSKEQAEVSISLYNKDIQDPVNFLRPVKLVKEERAAQPLKFKAIAAEQ
 170 180 190 X X 200 210 220 230
 TIQFDEFROIWEKSRLSDLEDGVQEKFYTNDASASKSTYTFVENAEPYLVHTAFRAGLAKVI
 240 250 260 270 280 290

6. CELSA-4 (1-5)

A30828 Steroid 17alpha-monooxygenase cytochrome P450 17 -

ENTRY A30828 #Type Protein
 TITLE Steroid 17alpha-monooxygenase cytochrome P450 17 -
 Rat #EC-number 1.14.99.9
 DATE 19-May-1989 #Sequence 19-May-1989 #Text 10-Aug-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 SOURCE Rattus norvegicus #Common-name Norway rat
 ACCESSION A30828\ A31359
 REFERENCE
 #Authors Dufau M.L.
 #Citation submitted to GenBank, December 1988.
 #Reference-number A94511
 #Accession A30828
 #Molecule-type mRNA
 #Residues 1-507 <DUF>

```

#Authors      NAMIKI M., KITAHARA M., BUCZKO E., DUTAU M.L.
#Journal      Biochem. Biophys. Res. Commun. (1988) 157:705-712
#Title        Rat testis P-450-17-alpha cDNA: the deduced amino
               acid sequence, expression and secondary structural
               configuration.
#Reference-number A90154
#Accession     A31359
#Molecule-type mRNA
#Residues      1-507 <NAM>
SUPERFAMILY    #Name cytochrome P450
KEYWORDS        endoplasmic reticulum\ heme\ membrane protein\
               monooxygenase\ oxidoreductase
FEATURE
  2-21          #Domain transmembrane <TM1>\
  169-186       #Domain transmembrane <TM2>\
  441          #Binding-site heme iron (Cys) (axial
               ligand) (predicted)
SUMMARY        #Molecular-weight 57250 #Length 507 #Checksum 9025
SEQUENCE
COMMENT        Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score   =      5   Optimized Score   =      5   Significance =   2.09
Residue Identity =   100%   Matches           =      5   Mismatches   =      0
Gaps            =      0   Conservative Substitutions =      0

```

```

ENEWDQPDQFMPEFLDPTGSHLITPTQSYLPFGAGPRSCIGEALARGELEVFETALLLRFDLDVSDDKQLP
  410          420          430          440          450          460          470
               X      X
               QAEVS
               I I I I
RLEGDPKVVFLLIDPFKVKITVRQAWMDAQAEVST
  480          490          500 X      X

```

7. CELSA-4 (1-5)

S16719 #Steroid 17alpha-monooxygenase - Rat #EC-number

```

ENTRY          S16719      #Type Protein
TITLE          #Steroid 17alpha-monooxygenase - Rat #EC-number
               1.14.99.9
DATE           15-Jun-1992 #Sequence 15-Jun-1992 #Text 15-Jun-1992
PLACEMENT      0.0    0.0    0.0    0.0    0.0
COMMENT        #This entry is not verified.
SOURCE         Rattus norvegicus #Common-name Norway rat
REFERENCE
  #Authors      Fevold H.R., Lorence M.C., McCarthy J.L., Trant
               J.M., Kagimoto M., Waterman M.R., Mason J.I.
  #Journal      Mol. Endocrinol. (1989) 3:968-975
  #Title        Rat P450(17-alpha) from testis: characterization of
               a full-length cDNA encoding a unique steroid
               hydroxylase capable of catalyzing both Delta(4)-
               and Delta(5)-steroid-17,20-lyase reactions.
  #Reference-number S16719
  #Accession     S16719
  #Cross-reference EMBL:M31681
SUMMARY        #Molecular-weight 57250 #Length 507 #Checksum 9025
SEQUENCE
COMMENT        Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score   =      5   Optimized Score   =      5   Significance =   2.09
Residue Identity =   100%   Matches           =      5   Mismatches   =      0
Gaps            =      0   Conservative Substitutions =      0

```

```

ENEWDQPDQFMPEFLDPTGSHLITPTQSYLPFGAGPRSCIGEALARGELEVFETALLLRFDLDVSDDKQLP

```

X X
QAEVS
|||||
RLEGDPKVVFLIDPFKVKITVRQAWMDAQAEVST
480 490 500 X X

8. CELSA-4 (1-5)

CPT1_RAT CYTOCHROME P450 XVIIIA1 (P450-C17) (EC 1.14.99.9) (

ID CPT1_RAT STANDARD; PRT; 507 AA.
AC P11715;
DT 01-OCT-1989 (REL. 12, CREATED)
DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
DT 01-MAR-1992 (REL. 21, LAST ANNOTATION UPDATE)
DE CYTOCHROME P450 XVIIIA1 (P450-C17) (EC 1.14.99.9) (STEROID 17-ALPHA-
DE HYDROXYLASE/17,20 LYASE).
GN CYP17.
OS RATTUS NORVEGICUS (RAT).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUTHERIA; RODENTIA.
RN [1]
RP SEQUENCE FROM N.A.
RM 89295447
RA FEVOLD H.R., LORENCE M.C., MCCARTHY J.L., TRANT J.M., KAGIMOTO M.,
RA WATERMAN M.R., MASON J.I.;
RL MOL. ENDOCRINOL. 3:968-975(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RM 89076306
RA NAMIKI M., KITAMURA M., BUCZKO E., DUFAU M.L.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 157:705-712(1988).
RN [3]
RP SEQUENCE OF 271-507 FROM N.A.
RM 88280759
RA NISHIHARA M., WINTERS C.A., BUCZKO E., WATERMAN M.R., DUFAU M.L.;
RL BIOCHEM. BIOPHYS. RES. COMMUN. 154:151-158(1988).
RN [4]
RP SEQUENCE OF 273-507 FROM N.A.
RM 90046678
RA MELLON S.H., VAISSE C.;
RL PROC. NATL. ACAD. SCI. U.S.A. 86:7775-7779(1989).
CC -!- FUNCTION: CYTOCHROMES P450 ARE A GROUP OF HEME-THIOLATE
CC MONOOXYGENASES. THEY OXIDIZE A VARIETY OF STRUCTURALLY UNRELATED
CC COMPOUNDS, INCLUDING STEROIDS, FATTY ACIDS, AND XENOBIOTICS.
CC -!- CATALYTIC ACTIVITY: A STEROID + AH(2) + O(2) = A 17-ALPHA-
CC HYDROXYSTEROID + A + H(2)O.
DR EMBL; M31681; RNP45017.
DR EMBL; M22204; RNP450X.
DR EMBL; M21208; RRCYPC17.
DR EMBL; M27282; RNP450C1.
DR PIR; A27659; A27659.
DR PIR; A30828; A30828.
DR PIR; S16719; S16719.
DR PROSITE; PS00086; CYTOCHROME_P450.
KW ELECTRON TRANSPORT; OXIDOREDUCTASE; MONOOXYGENASE; MEMBRANE;
KW HEME; STEROIDOGENESIS.
FT BINDING 441 441 HEME.
SQ SEQUENCE 507 AA; 57250 MW; 1347075 CN;
CC -!- Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
Residue Identity = 100% Matches = 5 Mismatches = 0

ENEWDQPDQFMFERFLDPTQSHLITPTQSYLPFCAGPRBQICEALARGELVFETALLORFDLDVSDDKQLP
410 420 430 440 450 460 470

X X
QAEVS
IIIII
RLEGDPKVVFLIDPFKVKITVRQAWMDAQAQEVST
480 490 500 X X

9. CELSA-4 (1-5)

JH0594

Vasoactive intestinal polypeptide receptor precursor

ENTRY JH0594 #Type Protein
TITLE Vasoactive intestinal polypeptide receptor precursor
- Rat
DATE 17-Jul-1992 #Sequence 17-Jul-1992 #Text 17-Jul-1992
PLACEMENT 0.0 0.0 0.0 0.0 0.0
SOURCE Rattus norvegicus #Common-name Norway rat
ACCESSION JH0594
REFERENCE (Lung)
#Authors Ishihara T., Shigemoto R., Mori K., Takahashi K.,
Nagata S.
#Journal Neuron (1992) 8:811-819
#Title Functional expression and tissue distribution of a
novel receptor for vasoactive intestinal
polypeptide.
#Reference-number JH0594
#Accession JH0594
#Molecule-type mRNA
#Residues 1-459 (ISH)
#Cross-reference GB:M86835

FEATURE

1-30

#Domain signal sequence (predicted)
<SIG>\

31-459

#Protein vasoactive intestinal
polypeptide receptor (predicted) <MAT>\
#Binding-site carbohydrate (Asn)
(covalent) (predicted)\

146-168

#Domain transmembrane <TM1>\

176-195

#Domain transmembrane <TM2>\

218-241

#Domain transmembrane <TM3>\

256-277

#Domain transmembrane <TM4>\

295-318

#Domain transmembrane <TM5>\

344-363

#Domain transmembrane <TM6>\

376-395

#Domain transmembrane <TM7>\

#Molecular-weight 52057 #Length 459 #Checksum 2598

SUMMARY

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

GIHYVNFAPFPDNRRAQVKMVFELVVGSGGQVVAIFYCFLNGEVDAGLRSFMRWHLGGVUQHSKSKOHFW
350 370 380 390 400 410 420

X X

QAEVS

IIIII

GGGHSATGTAEMLTVPDQARVSSSFQAEVSLV
430 440 450

--
nolF protein - Rhizobium meliloti

ENTRY S16562 #Type Protein
TITLE #nolF protein - Rhizobium meliloti
DATE 11-Jan-1992 #Sequence 11-Jan-1992 #Text 11-Jan-1992
PLACEMENT 0.0 0.0 0.0 0.0 0.0
COMMENT *This entry is not verified.
SOURCE Rhizobium meliloti
REFERENCE

#Authors Baev N., Endre G., Petrovics G., Banfalvi Z.,
Kondorosi A.
#Journal Mol. Gen. Genet. (1991) 228:113-124
#Title Six nodulation genes of nod box locus 4 in Rhizobium
meliloti are involved in nodulation signal
production: nodM codes for D-glucosamine
synthetase.

#Reference-number S16561

#Accession S16562

#Cross-reference EMBL:X58632

SUMMARY #Molecular-weight 39541 #Length 367 #Checksum 9824

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:42:00-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

LTPIRRSTLTSRVSRRLTSCSDRGRGQCRRSPPRFRKSSSRIGGDSAKSRGGCLSAQTELAEEAVLERNTRL
70 80 90 100 110 120 130

X X
QAEVS

GERGAASEATRLAALADVLDLRAHVRSKQAEVSDAERSLSHAEVRAEFGGVIRARSGVEEGQTVPLNTQLMTI
140 150 160 X 170 180 190 200

VELNRLEVDAGVPTSRIPILRLKQSVELMVEGFPGRTFSGEVARISPTADAGSRAVRVFIA
210 220 230 240 250 260

11. CELSA-4 (1-5)

NOLF_RHIME NODULATION PROTEIN NOLF.

ID NOLF_RHIME STANDARD; PRT: 367 AA.
AC P25196;
DT 01-MAY-1992 (REL. 22, CREATED)
DT 01-MAY-1992 (REL. 22, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE NODULATION PROTEIN NOLF.
GN NOLF.
OS RHIZOBIUM MELILOTI.
OG PLASMID SYM PRME41B.
OC PROKARYOTA; GRACILICUTES; SCOTOBACTERIA; AEROBIC RODS AND COCCI;
OC RHIZOBIACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AK 631;
RM 71360053
RA BAEV N., ENDRE G., PETROVICS G., BANFALVI Z., KONDOROSI A.
RL MOL. GEN. GENET. 228:113-124(1991).
CC 41- FUNCTION: INVOLVED IN THE PRODUCTION OF NODULATION-SPECIFIC
CC NODULATION SIGNAL MOLECULE.
DI EMBL: X58632; GDB:PRME41B.
DR
PR

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

LTPIRRSTLTSRVSRRLLTSCSDRGRGQCRRSPRPFRKGSRRIGGDGANSRGGCLSAOTELAEAVLERNTRL
 70 80 90 100 110 120 130

X X
 QAEVS

|||||

GERGAASEATRLAALADVLDLRAHVRSKQAEVSDAERSLSHAEVRAEFGGVIRARSVEEGQTVPLNTQLMTI
 140 150 160 170 180 190 200

VELNRLEVDAGVPTSRIPILRLKQSVELMVEGFPGRTFSGEVARISPTADAGSRAVRVFIA
 210 220 230 240 250 260

12. CELSA-4 (1-5)

A27659 Cytochrome P450 17 - Rat (fragment)

ENTRY A27659 #Type Protein (fragment)
 TITLE Cytochrome P450 17 - Rat (fragment)
 ALTERNATE-NAME cytochrome P450-17-alpha
 DATE 31-Mar-1989 #Sequence 31-Mar-1989 #Text 10-Aug-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 SOURCE Rattus norvegicus #Common-name Norway rat
 ACCESSION A27659
 REFERENCE

#Authors Nishihara M., Winters C.A., Buzko E., Waterman M.R.,
 Dufau M.L.

#Journal Biochem. Biophys. Res. Commun. (1988) 154:151-158

#Title Hormonal regulation of rat Leydig cell cytochrome
 P-450-17-alpha mRNA levels and characterization of
 a partial length rat P-450-17-alpha cDNA.

#Reference-number A27659

#Accession A27659

#Molecule-type mRNA

#Residues 1-237 <NIS>

#Comment The authors translated the codon GAT for residues
 18, 131, and 208 as Glu, Asn, and Asn
 respectively.

SUPERFAMILY #Name cytochrome P450

KEYWORDS heme\ monooxygenase\ oxidoreductase

FEATURE

171 #Binding-site heme iron (Cys) (axial
 ligand).

SUMMARY

SEQUENCE #Length 237 #Checksum 8451

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:42:01-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

ENEWDQPDQFMPEFLDPTCSHLITPTQSYLPFCAGPRSCIGEALAQSELPFTALLERFELDVSDDKOLF
 140 150 160 170 180 190 200

X X
 QAEVS

|||||

RLADPRVUVFLIDPRKUYITIRGANIDAGAEVRS

112

121

200

A33980

Steroid 17alpha-monooxygenase cytochrome P450 17 -

ENTRY
TITLEA33980 *Type Protein (fragment)
Steroid 17alpha-monooxygenase cytochrome P450 17 -
Rat (fragment) *EC-number 1.14.99.9DATE
PLACEMENT
SOURCE31-Mar-1992 *Sequence 31-Mar-1992 *Text 10-Aug-1992
0.0 0.0 0.0 0.0 0.0ACCESSION
REFERENCERattus norvegicus *Common-name Norway rat
A33980#Authors
#Journal
#TitleMellon S.H., Vaisse C.
Proc. Natl. Acad. Sci. U.S.A. (1989) 86:7775-7779
cAMP regulates P450scc gene expression by a
cycloheximide-insensitive mechanism in cultured
mouse Leydig MA-10 cells.

#Reference-number A33980

#Accession A33980

#Molecule-type mRNA

#Residues 1-235 <MEL>

SUPERFAMILY

#Name cytochrome P450

KEYWORDS

heme\ monooxygenase\ oxidoreductase

FEATURE

169

#Binding-site heme iron (Cys) (axial
ligand) (predicted)

SUMMARY

SEQUENCE

#Length 235 #Checksum 7248

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:42:01-PST using FastDB

Initial Score	=	5	Optimized Score	=	5	Significance	=	2.09
Residue Identity	=	100%	Matches	=	5	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

ENEWDQPDQFMPEFLDPTGSHLITPTOSYLPFGAGPRSCIGEALARGELFVFTALLQRFDLVDVSDDKQLP
 130 140 150 160 170 180 190 200

X X

QAEVS

IIIII

RLEGDPKVVFLLIDPFKVKITVRQAWMDAQAEVST
 210 220 230 X

14. CELSA-4 (1-5)

S18659

*Hypothetical protein - Mycoplasma hyorhinis (SGC3)

ENTRY

S18659 *Type Protein

TITLE

*Hypothetical protein - Mycoplasma hyorhinis (SGC3)

DATE

02-Sep-1992 *Sequence 02-Sep-1992 *Text 02-Sep-1992

PLACEMENT

0.0 0.0 0.0 0.0 0.0

COMMENT

*This entry is not verified.

SOURCE

Mycoplasma hyorhinis

REFERENCE

#Authors

Yogev D., Rosengarten R., Watson-Mckown R., Wise
R.R.

#Journal

EMBO J. (1991) 10:4069-4079

#Title

Molecular basis of Mycoplasma surface antigen
variation: a comparison of the 5' and 3' regions
spontaneous mutation of the 5' and 3' regions
and 5' regulatory sequence.

#Reference-number S18659

#Accession S18659

#Sequence-reference EMBL: A0336

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:42:01-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

```

          X   X
          QAEVS
          |||||
VDKLIQINHHNNQDLVHLOVEINLEQQQVDIQDQEHLOAEVSLQPVVLEQQQVVKQKLLLNQNDHKKOLL
      10      20      30      X 40      50      60      70
NLQNDHKKOLLNLQNDHKKOLLNLQNDHKKOLLNLQNDHKKOLLNLQNDHKKLLKKKLQTKNPI
      80      90     100     110     120     130     140

```

15. CELSA-4 (1-5)

R20793 CDR-grafted, humanised heavy chain gH1.

ID R20793 standard; Protein: 146 AA.
 AC R20793;
 DT 19-MAY-1992 (first entry)
 DE CDR-grafted, humanised heavy chain gH1.
 KW murine monoclonal antibody; MAb; A5B7; humanised antibody; CEA;
 KW complementarity determining region.
 OS Homo sapiens.
 OS Mus musculus.
 FH Key Location/Qualifiers
 FT Peptide 1..19
 FT /label= signal
 FT Protein 20..146
 FT /label= VH
 FT /note= "human LAY framework with A5B7 CDRs"
 FT Region 45..54
 FT /label= CDR1
 FT /note= "murine residues"
 FT Region 69..87
 FT /label= CDR2
 FT /note= "murine residues"
 FT Region 120..129
 FT /label= CDR3
 FT /note= "murine residues"
 FT Misc_difference 20
 FT /note= "murine residue"
 FT Misc_difference 67..68
 FT /note= "murine residues"
 FT Misc_difference 94..95
 FT /note= "murine residues"
 FT Misc_difference 98
 FT /note= "murine residue"
 FT Misc_difference 118
 FT /note= "murine residue"
 PN WD9201059-A.
 PD 23-JAN-1992.
 PF 05-JUL-1991; G01108.
 PR 05-JUL-1990; GB-014932.
 PR 21-DEC-1990; WD-G02017.
 PR 05-JUL-1991; WD-G01108.
 PA (CELL-) CELLTECH LTD.
 PI Adair JR, Bodmer MW, Mountain A, Owens RJ;
 DR WPI; 92-056874/07.
 DR N-P80B; Q20987.
 PT New CDR-grafted anti-mouse monoclonal antibody; A5B7; CEA;
 FT is thereby humanised by grafting the murine CDRs onto the human

CC heavy chain sequence comprises a human framework (i.e. the LAY
 CC region) which contains murine sequences (from the murine anti-CEA
 CC A5B7 MAb) in the CDRs and at other positions predicted to be
 CC important for antigen-binding and at which human and A5B7 sequences
 CC differ. (See Q20984 for A5B7 heavy chain coding sequence).
 SQ Sequence 146 AA;
 SQ 7 A: 7 R: 4 N: 4 D: 0 B: 2 C: 6 Q: 6 E: 0 Z: 18 G: 1 H:
 SQ 3 I: 14 L: 6 K: 3 M: 9 F: 3 P: 16 S: 13 T: 5 W: 9 Y: 10 V:
 CC Retrieved by alexk on Thu 25 Feb 93 10:42:01-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

LFFLSVTTGVHSEVQLLESGLLVQPGGSLRLSCATSGFTFTDYIMNWVRQAPGKGLEWLGFIGNKANGYTT
 10 20 30 40 50 60 70

X X

QAEVS

|||||

EYSASVKGRFTISRDKSKSTLYLQMNGLQAEVSAIYYCTDRGLRFYFDYWGGTGLVTVSSASTKGP
 80 90 100 110 X 120 130 140

16. CELSA-4 (1-5)

S06727

*Hypothetical protein 1 (minicircle) - Streptomyce

ENTRY S06727 #Type Protein
 TITLE *Hypothetical protein 1 (minicircle) - Streptomyces
 coelicolor

DATE 15-Jun-1992 #Sequence 15-Jun-1992 #Text 15-Jun-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0

COMMENT *This entry is not verified.

SOURCE Streptomyces coelicolor

REFERENCE

#Authors Henderson D.J., Lydiate D.J., Hopwood D.A.
 #Journal Mol. Microbiol. (1989) 3:1307-1318
 #Title Structural and functional analysis of the
 mini-circle, a transposable element of
 Streptomyces coelicolor A3(2).

*Reference-number S06727

*Accession S06727

*Cross-reference EMBL:X15942

SUMMARY #Molecular-weight 13374 #Length 122 #Checksum 9249
 SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:42:01-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 2.09
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

X X

QAEVS

|||||

MTLGLVQHNAEVERNWVFQRFVAGQNVPPVFGESENHGGFALKSGRGLDEAVAAWQAEVSRGREL IADAGLDD
 10 20 30 40 50 60 70

SGHLSEGEAGHVGDQGVSLRWIMVHMIEEYARHNGHADLIREQIDGTICA
 80 90 100 110 120

17. CELSA-4 (1-5)

YN2_S1R00 MINI-CIRCLE HYPOTHETICAL 13.3 K0 1000000

17 YN2_S1R00

```
Initial Score      =      5   Optimized Score   =      5   Significance =    2.09
Residue Identity  =    100%   Matches         =      5   Mismatches  =      0
Gaps              =      0   Conservative Substitutions =      0
```

```

X      X
QAEVS
|||||
MTLLGLVQHMAEVERNWFQRVFAQONVPPVFGESENHDFALKSGRCLDEAVAAWQAEVSRGRELIADAGLDD
  10          20          30          40          50          X      60          70
SGHLSEQEAQHVGDQGVSLRWIMVHMIEEYARHNGHADLIREQIDGTTGA
  80          90          100         110         120

```

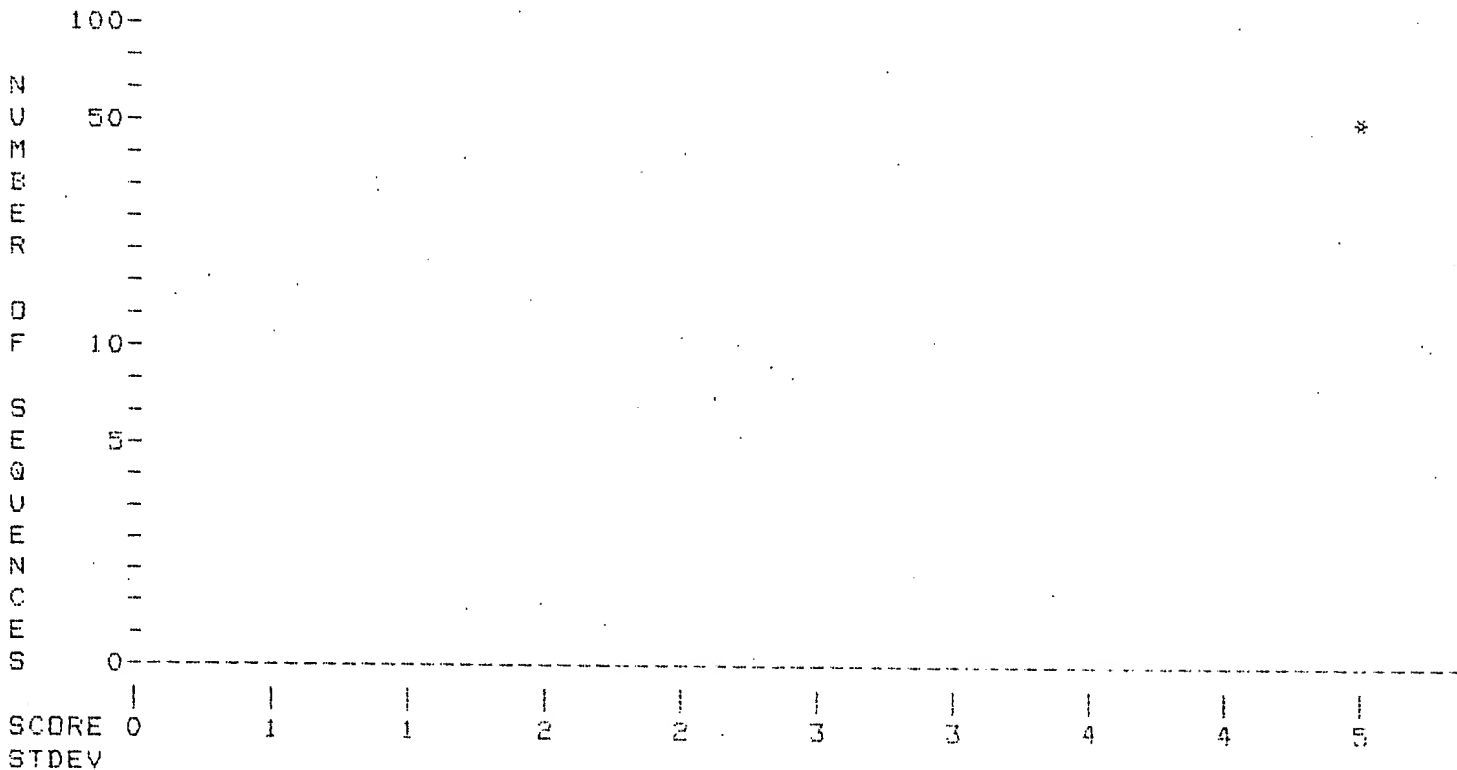
> Q <
 01 10 IntelliGenetics
 > Q <

FastDB - Fast Pairwise Comparison of Sequences
 Release 5.4

Results file celsa-14-align.res made by alexk on Thu 25 Feb 93 11:15:30-PST.

Query sequence being compared: CELSA-14 (1-5)
 Number of sequences searched: 50
 Number of scores above cutoff: 50

Results of the initial comparison of CELSA-14 (1-5) with:
 File : celsa-14.pep



PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	5	6	0.00

Times:	CPU	Total Elapsed
	00:00:01.03	00:00:01.05

Number of residues:	45939
Number of sequences searched:	50
Number of scores above cutoff:	50

50 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. S18268	*alpha-Aminoadipyl-L-cysteiny	3649	5	5	0.00	0
2. ACVS_NOCLA	L-(ALPHA-AMINOADIPYL)-L-CYSTE	3649	5	5	0.00	0
3. S12332	Ubiquitin--protein ligase - Y	1950	5	5	0.00	0
4. UBR1_YEAST	N-END-RECOGNIZING PROTEIN (UB	1950	5	5	0.00	0
5. RRWGNV	RNA-directed RNA polymerase -	1643	5	5	0.00	0
6. VOR1_NMV	186 KD PROTEIN (ORF 1).	1643	5	5	0.00	0
7. WMWGPV	RNA-directed RNA polymerase -	1456	5	5	0.00	0
8. S14005	*Hypothetical protein, 166K -	1456	5	5	0.00	0
9. VOR1_PVX	165 KD PROTEIN (ORF 1).	1456	5	5	0.00	0
10. VOR1_PVXCP	165 KD PROTEIN (ORF 1).	1456	5	5	0.00	0
11. VOR1_PVXX3	165 KD PROTEIN (ORF 1).	1456	5	5	0.00	0
12. VGIHJ2	E2 glycoprotein precursor - M	1376	5	5	0.00	0
13. VGL2_CVM4	E2 GLYCOPROTEIN PRECURSOR (SP	1376	5	5	0.00	0
14. VGIH59	E2 glycoprotein precursor - M	1324	5	5	0.00	0
15. VGL2_CVMA5	E2 GLYCOPROTEIN PRECURSOR (SP	1324	5	5	0.00	0
16. VGIHMJ	E2 glycoprotein precursor - M	1235	5	5	0.00	0
17. VGL2_CVMJH	E2 GLYCOPROTEIN PRECURSOR (SP	1235	5	5	0.00	0
18. A40986	*M-cadherin - Mouse (fragment	730	5	5	0.00	0
19. SAHU4F	Cell surface antigen 4F2 heav	529	5	5	0.00	0
20. A39984	*Cell surface antigen 4F2 hea	529	5	5	0.00	0

1. CELSA-14 (1-5)

S18268 *alpha-Aminoadipyl-L-cysteinyL-D-valine synthetase

ENTRY S18268 #Type Protein

TITLE *alpha-Aminoadipyl-L-cysteinyL-D-valine synthetase - Streptomyces lactamurans

DATE 02-Apr-1992 #Sequence 02-Apr-1992 #Text 02-Apr-1992

PLACEMENT 0.0 0.0 0.0 0.0 0.0

COMMENT *This entry is not verified.

SOURCE Streptomyces lactamurans

REFERENCE

*Authors Martin J.F.

*Citation submitted to the EMBL Data Library, January 1991

*Reference-number S18268

*Accession S18268

*Cross-reference EMBL:X57310

SUMMARY #Molecular-weight 404084 #Length 3649 #Checksum 2296

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:29:01-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00

Residue Identity = 100% Matches = 5 Mismatches = 0

Gaps = 0 Conservative Substitutions = 0

ALSSRDRAVVTYTSGITGVPAQVPRKYHYSVVMISITDLSEYDMRRPGTERVALFASYVFEPHLRQTLIALIN

400 410 420 430 440 450 460

X X

YLNAT

|||||

EGTLVIYPDDVRLDFDLFPEIERHGVTYLNATGSVLWHPDIKALASLARIACCTVIAKQVQIARFAL

470 480 490 500 510 520 530

RWVNEAATGAAAV... (truncated)

Best Available Copy

2. CELSA-14 (1-5)

ACVS_NOCLA L-(ALPHA-AMINOADIPYL)-L-CYSTEINYL-D-VALINE SYNTHET

ID ACVS_NOCLA STANDARD; PRT: 3649 AA.
AC P27743;
DT 01-AUG-1992 (REL. 23, CREATED)
DT 01-AUG-1992 (REL. 23, LAST SEQUENCE UPDATE)
DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
DE L-(ALPHA-AMINOADIPYL)-L-CYSTEINYL-D-VALINE SYNTHETASE (EC 6.-.-.-)
DE (ACV SYNTHETASE) (ACVS).
GN PCBAB.
OS NOCARDIA LACTAMDURANS.
OC PROKARYOTA; FIRMICUTES; ACTINOMYCETALES; NOCARDIOFORM.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VAR LC 411;
RM 92065808
RA COQUE J.J.R., MARTIN J.F., CALZADA J.G., LIRAS P.;
RL MOL. MICROBIOL. 5:1125-1133(1991).
CC -!- FUNCTION: EACH OF THE CONSTITUENT AMINO ACIDS OF ACV ARE ACTIVATED
CC AS AMIONACYL-ADENYLATES WITH PEPTIDE BONDS FORMED THROUGH THE
CC PARTICIPATION OF AMINO ACID THIOLESTER INTERMEDIATES.
CC -!- PATHWAY: FIRST STEP IN THE BIOSYNTHESIS OF PENICILLIN AND
CC CEPHALOSPORIN.
CC -!- SIMILARITY: TO OTHER ENZYMES WHICH ACT VIA AN ATP-DEPENDENT
CC COVALENT BINDING OF AMP TO THEIR SUBSTRATE.
DR EMBL: X57310; NLPBABC.
DR PIR: S18268; S18268.
KW LIGASE; ANTIBIOTIC BIOSYNTHESIS; DUPLICATION.
SQ SEQUENCE 3649 AA; 404079 MW; 2.019611E+07 CN;
CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score	=	5	Optimized Score	=	5	Significance	=	0.00
Residue Identity	=	100%	Matches	=	5	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

ALSSRDRAVVTYTSGTTGVPGKGVPKYHYSVNSITDLSEYDMRRPGTERVALFASYVFEPHLRQTLIALIN
400 410 420 430 440 450 460

X X
YLNAT
|||||

EGTLVIVPDDVRLDFDLFPEYIERHGVTYLNATGSVLQHFBLRRCASLKRLLLVGEELTASGLRQLREKFAQ
470 480 490 500 X 510 520 530 540

RVVNEYAFTEAAFTAVKEFGPGVTERRDRSIGRPLRNWKYVLSQGLKQLPFGAIGELYI
550 560 570 580 590 600

3. CELSA-14 (1-5)

S12332

Ubiquitin--protein ligase - Yeast (*Saccharomyces*

ENTRY S12332 #Type Protein
TITLE Ubiquitin--protein ligase - Yeast (*Saccharomyces cerevisiae*) #EC-number 6.3.2.19
DATE 16-Sep-1992 #Sequence 16-Sep-1992 #Text 29-Sep-1992
PLACEMENT 0.0 0.0 0.0 0.0 0.0
SOURCE *Saccharomyces cerevisiae*
ACCESSION S12332
REFERENCE

#Author Bartel P., Unanue J., Vassilakis A
#Journal EMBO J. (1990) 9:3177-3187
#Title The Yeast Ubiquitin-Protein Ligase
#Accession 12332

#Residues 1-1950 (BAR)
#Cross-reference EMBL:X53747

GENETIC

#Map-position 15

#Name UBR1

SUPERFAMILY #Name ubiquitin--protein ligase
KEYWORDS protein degradation

SUMMARY #Molecular-weight 224836 #Length 1950 #Checksum 7669

SEQUENCE
COMMENT Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

ATTALRQGVPDNVHIDLLTSRIDGEGRAMLKCSQDLSSVLGGFFAVQTNGLSATLTWSSEYLHQETCKYIIL
330 340 350 360 370 380 390

X X

YLNAT

IIII

WITHCLNIPNSSFQTTFRNMMGKTLCEYLNATECRDMTPVVEKYFSNKFCKNDPYRYIDLSILADGNQIFL
400 410 420 X 430 440 450 460

GHHKILPESSTHSLSPINDVETPTSRYSNTRLQHILYFDNRYWKRLRKDIQNVIIPTLA
470 480 490 500 510 520 530

4. CELSA-14 (1-5)

UBR1_YEAST N-END-RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGAS

ID UBR1_YEAST STANDARD; PRT; 1950 AA.
AC P19812;
DT 01-FEB-1991 (REL. 17, CREATED)
DT 01-FEB-1991 (REL. 17, LAST SEQUENCE UPDATE)
DT 01-AUG-1991 (REL. 19, LAST ANNOTATION UPDATE)
DE N-END-RECOGNIZING PROTEIN (UBIQUITIN-PROTEIN LIGASE E3 COMPONENT) (N-
DE RECOGNIN).
GN UBR1.
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).
OC EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOMYCETES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GRF88 (S288C);
RM 91006011
RA BARTEL B., WUENNING I., VARSHAVSKY A.;
RL EMBO J. 9:3179-3189(1990).
CC -!- FUNCTION: UBR1 IS THE RECOGNITION COMPONENT OF THE N-END RULE
CC PATHWAY. UBR1 BINDS TO PROTEINS BEARING AMINO-TERMINAL RESIDUES
CC THAT ARE DESTABILIZING ACCORDING TO THE N-END RULE, BUT DOES NOT
CC BIND TO OTHERWISE IDENTICAL PROTEINS BEARING STABILIZING AMINO-
CC TERMINAL RESIDUES.
DR EMBL: X53747; SCUBRIG.
DR PIR: S12332; S12332.
KW UBIQUITIN CONJUGATION.
SQ SEQUENCE 1950 AA; 224836 MW; 1.833403E+07 CN;
CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02 PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

ATTALRQGVPDNVHIDLLTSRIDGEGRAMLKCSQDLSSVLGGFFAVQTNGLSATLTWSSEYLHQETCKYIIL
330 340 350 360 370 380 390

Best Available Copy X

YLNAT

|||||

WITHCLNIPNSSFOTTFRNMMGKTLCEYLNATECRDMTPVVEKYFSNKFDDKNDPYRYIDLSILADGNQIFL
400 410 420 X 430 440 450 460

GHHKILPESSTHSLSPILNDVETPTSRITYGNTRLQHILYFDNRYWKRLRKDIONVVIPTLA
470 480 490 500 510 520 530

5. CELSA-14 (1-5)

RRWGNV

RNA-directed RNA polymerase - Narcissus mosaic vir

ENTRY

TITLE

RRWGNV

#Type Protein

RNA-directed RNA polymerase - Narcissus mosaic virus
#EC-number 2.7.7.48

ALTERNATE-NAME

RNA replicase

DATE

30-Sep-1992 #Sequence 30-Sep-1992 #Text 30-Sep-1992

PLACEMENT

364.0 3.0 1.0 1.0 1.0

SOURCE

narcissus mosaic virus

ACCESSION

JT0470

REFERENCE

#Authors

Zuidema D., Linthorst H.J.M., Huisman M.J., Asjes
C.J., Bol J.F.

#Journal

J. Gen. Virol. (1989) 70:267-276

#Title

Nucleotide sequence of narcissus mosaic virus RNA.

#Reference-number JT0470

#Accession

JT0470

#Molecule-type genomic RNA

#Residues

1-1643 <ZUI>

#Cross-reference GB:D00405

COMMENT

This virus is a member of the Potexvirus group.

SUPERFAMILY

#Name eggplant mosaic virus RNA-directed RNA
polymerase

KEYWORDS

nucleotidyltransferase

SUMMARY

SEQUENCE

COMMENT

#Molecular-weight 186303 #Length 1643 #Checksum 8014

Retrieved by alexk on Thu 25 Feb 93 10:27:02-PST using FastDB.

Initial Score =

=

5

Optimized Score =

=

5

Significance = 0.00

Residue Identity =

=

100%

Matches =

=

5

Mismatches = 0

Gaps =

=

0

Conservative Substitutions =

= 0

VTPTVLLRNDWQTKLPILPADVFKTFEKSIVIQPCNPILVFDDYTKLPGLIESVVMHHQNVVFIIILTGDNRRQ
900 910 920 930 940 950 960

X X

YLNAT

|||||

SVYHETNPEAYIAALPEAVEIFSPYCEFYLNATHRNVDLANKLGVYSERECKLVNFGASHHLKASRIEMLV
970 980 990 X 1000 1010 1020 1030

PSTMKRNAMFDMGHHSMTYAGCQGLTAPKIGILLDNHTQFCSESVLYTCLERAVDRIHF IN
1040 1050 1060 1070 1080 1090

5. CELSA-14 (1-5)

VOR1_NMV

185 KD PROTEIN (ORF 12)

ID

VOR1_NMV

STANDARD:

PROT: 1643 AA.

AC

P15098

DT

01-APR-1990 (REL. 14, CREATED)

DT

01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)

DT

01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)

TF

126 15 1000000 1000000

RN [1]
 RP SEQUENCE FROM N.A.
 RM 89279206
 RA ZUIDEMA D., LINTHORST H.J.M., HUISMAN M.J., ASJES C.J., BOL J.F.,
 RL J. GEN. VIROL. 70:267-276(1989).
 CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
 CC POSSIBLY FUNCTIONS AS A NTP-BINDING HELICASE.
 DR EMBL: D00405; NMV.
 DR PIR: JT0470; JT0470.
 KW NUCLEOTIDE-BINDING; HELICASE; RNA REPLICATION.
 SQ SEQUENCE 1643 AA; 186304 MW; 1.392861E+07 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VTPTVLLRNDWQTKLPILPADVFKTFEKSVIOPCNFILVFDDYTKLPPGLIESVVMHHQNVVFIILTGDNRQ
 900 910 920 930 940 950 960

X X
 YLNAT

SVYHETNPEAYIAALPEAVEIFSPYCEFYLNATHRNVDLANKLGVYSEREGLKLVNFASHHLKASRIPMLV
 970 980 990 X 1000 1010 1020 1030

PSTMKRNAMEFDMGHHSMTYAGCGLTAPKIQILLDNHTQFCSESVLYTCLSRVADRIHFIN
 1040 1050 1060 1070 1080 1090

7. CELSA-14 (1-5)

WMWGPV RNA-directed RNA polymerase - Potato virus X (stra

ENTRY WMWGPV #Type Protein
 TITLE RNA-directed RNA polymerase - Potato virus X (strain
 X3) #EC-number 2.7.7.48
 ALTERNATE-NAME RNA replicase
 DATE 30-Jun-1990 #Sequence 30-Jun-1990 #Text 30-Sep-1992
 PLACEMENT 364.0 4.0 1.0 1.0 1.0
 SOURCE potato virus X, PVX
 HOST Nicotiana tabacum cv. Samsun #Common-name tobacco
 ACCESSION JA0102
 REFERENCE

#Authors Huisman M.J., Linthorst H.J.M., Bol J.F.,
 Cornelissen B.J.C.
 #Journal J. Gen. Virol. (1988) 69:1789-1798
 #Title The complete nucleotide sequence of potato virus X
 and its homologues at the amino acid level with
 various plus-stranded RNA viruses.

#Reference-number JA0102
 #Accession JA0102
 #Molecule-type mRNA
 #Residues 1-1456 (HUI)
 #Cross-reference GB:M31541

COMMENT This virus is a member of the Potexvirus group.
 SUPERFAMILY #Name eggplant mosaic virus RNA-directed RNA
 polymerase
 KEYWORDS nucleotidyltransferase
 SUMMARY #Molecular-weight 155505 #Length 1456 #Accession 6078
 SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VLPTNELRLDWSKKVPNTEPYMFKTYEKALIGGTGSIVIFDDYSKLPFGYIEALVCFYSKIKLIILTGDSRQ
 770 780 790 800 810 820 830

X X

YLNAT

IIIII

SVYHETAEDASIRHLGPATEYFYSKYCRYYLNATHRNKKDLANMLGVYSERTGVTEISMSAEFLEGIPTLVPS
 840 850 860 X 870 880 890 900

DEKRRKLYMGTRNDTFTYAGCQGLTKPKVQIVLDHNTQVCSANVMYTALSRATDRIHFVNT
 910 920 930 940 950 960

8. CELSA-14 (1-5)

S14005

*Hypothetical protein, 166K - Potato virus X

ENTRY S14005 #Type Protein
 TITLE *Hypothetical protein, 166K - Potato virus X
 DATE 16-Apr-1992 #Sequence 16-Apr-1992 #Text 16-Apr-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT *This entry is not verified.
 SOURCE potato virus X, PVX
 REFERENCE

#Authors Orman B.E., Celnik R.M., Mandel A.N., Torres H.N.,
 Mentaberry A.N.

#Journal Virus Res. (1990) 16:293-306

#Title Complete cDNA sequence of a South American isolate
 of potato virus X.

#Reference-number S14005

#Accession S14005

#Cross-reference EMBL:X55802

SUMMARY #Molecular-weight 165300 #Length 1456 #Checksum 9959

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VLPTNELRLDWSKKVPNTEPYMFKTYEKALIGGTGSIVIFDDYSKLPFGYIEALVSFSTKIKLIILTGDSRQ
 770 780 790 800 810 820 830

X X

YLNAT

IIIII

SVYHETSDDASIRHLGPATEVFAKYCRYYLNATHRNKKDLANMLGVYSERTGTTTEISMSSEFLEGVPTLVPS
 840 850 860 X 870 880 890 900

DEKRRRLYMGTRNDTFTYAGCQGLTKPKVQIVLDHNTQVCSANVMYSALSRATDRIHFINT
 910 920 930 940 950 960

9. CELSA-14 (1-5)

VOR1_PVX

165 KD PROTEIN (ORF 1).

ID VOR1_PVX STANDARD; PRT; 1456 AA.
 AC P09395;
 DT 01-MAR-1989 (REL. 10, CREATED)
 DT 01-MAR-1989 (REL. 10, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE 165 KD PROTEIN (ORF 1).
 OS POTATO VIRUS X (P. X).
 OC VIRIDAE: GB-RNA: COMPEL: 165 KD PROTEIN (ORF 1).
 RN 1

Best Available Copy

RN 87083320
 RA SKRYABIN K.G., KRAEV A.B., MURZOV S.Y., ROZANOV M.N., CHERNOV B.K.,
 RA LUKASHEVA L.I., ATABEKOV J.G.;
 RL NUCLEIC ACIDS RES. 16:10929-10930(1988).
 CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
 CC POSSIBLY FUNCTIONS AS A NTP-BINDING HELICASE.
 DR EMBL: X05198; POPVX3.
 KW NUCLEOTIDE-BINDING; HELICASE; RNA REPLICATION.
 FT NP_BIND 735 742 POTENTIAL.
 SQ SEQUENCE 1456 AA; 165406 MW; 1.09471E+07 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VLPTNELRLDWSKKVPNTEPYMFKTYEKALIGGTGSIVIFDDYSKLPFGYIEALICFYISKIKLVILTGDSSRQ
 770 780 790 800 810 820 830

X X
 YLNAT
 I I I I
 SVYHETAEDASIRHLGPATEYFSKYCRYLYNATHRNKKDLANMLGVYSGERTGVTEISMSAEFLEGIPTLVPS
 840 850 860 X 870 880 890 900

DEKRRLYMGTCGRNDTFTYAGCGGLTKPKVQIVLDHNTQVCSANVMTALSPATDRIHFVNT
 910 920 930 940 950 960

10. CELSA-14 (1-5)

VOR1_PVXCP 165 KD PROTEIN (ORF 1).

ID VOR1_PVXCP STANDARD; PRT; 1456 AA.
 AC P22591;
 DT 01-AUG-1991 (REL. 19, CREATED)
 DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE 165 KD PROTEIN (ORF 1).
 OS POTATO VIRUS X (STRAIN CP) (PVX).
 OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; POTEXVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 90364772
 RA ORMAN B.E., CELNIK R.M., MANDEL A.M., TORRES H.N., MENTABERRY A.N.;
 RL VIRUS RES. 16:293-306(1990).
 CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
 CC POSSIBLY FUNCTIONS AS A NTP-BINDING HELICASE.
 DR EMBL: X55802; POPVX.
 DR PIR: S14005; S14005.
 KW NUCLEOTIDE-BINDING; HELICASE; RNA REPLICATION.
 FT NP_BIND 735 742 POTENTIAL.
 SQ SEQUENCE 1456 AA; 165301 MW; 1.094681E+07 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VLPTNELRLDWSKKVPNTEPYMFKTYEKALIGGTGSIVIFDDYSKLPFGYIEALICFYISKIKLVILTGDSSRQ
 770 780 790 800 810 820 830

X X
 YLNAT

11. CELSA-14 (1-5)

VOR1_PVXX3 165 KD PROTEIN (ORF 1).

ID VOR1_PVXX3 STANDARD; PRT: 1456 AA.
 AC P17779;
 DT 01-AUG-1990 (REL. 15, CREATED)
 DT 01-AUG-1990 (REL. 15, LAST SEQUENCE UPDATE)
 DT 01-AUG-1992 (REL. 23, LAST ANNOTATION UPDATE)
 DE 165 KD PROTEIN (ORF 1).
 OS POTATO VIRUS X (STRAIN X3) (PVX).
 OC VIRIDAE; SS-RNA NONENVELOPED VIRUSES; POTEXVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 88299944
 RA HUISMAN M.J., LINTHORST H.J.M., BOL J.F., CORNELISSEN B.J.C.;
 RL J. GEN. VIROL. 69:1789-1798(1988).
 CC -!- FUNCTION: RNA-REPLICATION. THE CENTRAL PART OF THIS PROTEIN
 CC POSSIBLY FUNCTIONS AS A NTP-BINDING HELICASE.
 DR EMBL: D00344; PVXX3.
 DR PIR: JAO102; WMWGPV.
 KW NUCLEOTIDE-BINDING; HELICASE; RNA REPLICATION.
 FT NP_BIND 735 742 POTENTIAL.
 SQ SEQUENCE 1456 AA; 165605 MW; 1.100132E+07 CN;
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

VLPTNELRLDWSKKVPNTEPYMFKTYEKALIGSTGSIVIFDDYSKLPPGYIEALVCFYISKIKLIILTGDSTRQ
 770 780 790 800 810 820 830

X X
 YLNAT
 I I I I I

SVYHETAEDASIRHLGPATEYFSKYCRYL NATHRNKKDLANMLGVYSERTGVTEISMSAEFLEGIPTLVPS
 840 850 860 X 870 880 890 900

DEKRRLYMGTGRNDTFTYAGCGLTKPKVQIVLDHNTQVCSANVMYTALSRATDRIHFVNT
 910 920 930 940 950 960

12. CELSA-14 (1-5)

VGIHJ2 E2 glycoprotein precursor - Murine hepatitis virus

ENTRY VGIHJ2 #Type Protein
 TITLE E2 glycoprotein precursor - Murine hepatitis virus
 (strain wild type MHV-4)
 ALTERNATE-NAME peplomer glycoprotein\ spike glycoprotein
 INCLUDES 90B glycoprotein\ 90A glycoprotein
 DATE 31-Mar-1991 #Sequence 31-Mar-1991 11:01:30 31-Mar-1991
 PLACEMENT 2231.0 4.0 1.0 3.0 1.0
 SOURCE murine hepatitis virus, MHV
 ACCESSION A33743
 REFERENCE
 #Authors Parker S.E., Gallagher T.D., Buchholz J.A.
 #Journal Virology (1987) 123:560-571
 #Title Sequence analysis reveals extensive homology
 between the E2 glycoprotein precursor and the
 E1 glycoprotein of murine hepatitis virus

FASTA
#Molecule-type genomic RNA
#Residues 1-1376 (PAR)
COMMENT This virus is a member of the family Coronaviridae.
SUPERFAMILY #Name coronavirus E2 glycoprotein
KEYWORDS glycoprotein\ membrane protein
FEATURE

1-14 #Domain signal sequence (SIG)\
15-1376 #Protein E2 glycoprotein (E2G)\
15-769 #Protein 90B glycoprotein (EGB)\
770-1376 #Protein 90A glycoprotein (EGA)\
1321-1338 #Domain transmembrane (TMN)\
31,60,134,192,357,435,
442,582,677,709,717,
740,789,806,896,945,
1178,1232,1242,1261,
1277,1298,1370

#Binding-site carbohydrate (Asn)
(covalent) (predicted)

SUMMARY #Molecular-weight 151881 #Length 1376 #Checksum 4481

SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

MLFVFILLLPSCLG YIGDFRCIQTVNYNGNNASAPSISTEAVDVSKGLGTYYYVLDRVYLNATLLLTGYYPVD
10 20 30 40 50 60 X 70
GSNYRNLALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNTPTGATSYEFTIVIGSLFGNTSYTVVLEPY
80 90 100 110 120 130 140
NNIIMASVCTYTICQLPY
150 160

13. CELSA-14 (1-5)

VGL2_CVM4 E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PE

ID VGL2_CVM4 STANDARD; PRT: 1376 AA.
AC P22432;
DT 01-AUG-1991 (REL. 19, CREATED)
DT 01-AUG-1991 (REL. 19, LAST SEQUENCE UPDATE)
DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLMER PROTEIN).
GN S.
OS MURINE CORONAVIRUS MHV (STRAIN WILD TYPE 4) (MHV-4).
OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; CORONAVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RM 90085815
RA PARKER S.E., GALLAGHER T.M., BUCHMEIER M.J.;
RL VIROLOGY 173:664-673(1989).
CC -!- FUNCTION: THE PEPLMER PROTEIN MEDIATES THE BINDING OF VIRIONS
CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
CC -!- SIMILARITY: NEARLY IDENTICAL TO THE E2 GLYCOPROTEINS FROM MHV-JHM
CC AND MHV-AS9 STRAINS, EXCEPT FOR AN N-TERMINAL INSERTION.
CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE GLYCOPROTEIN.
DR EMBL: M32789; MHVE2GLY.
DR PIR: A33748; VGLHJ2.
KW GLYCOPROTEIN, ENVELOPE PROTEIN, TRANSMEMBRANE, SIGNAL
FT SIGNAL

FT	CHAIN	770	1378	SPIKE PROTEIN S1 (90B).
FT	DOMAIN	15	1320	SPIKE PROTEIN S2 (90A).
FT	TRANSMEM	1321	1338	EXTRACELLULAR (POTENTIAL).
FT	DOMAIN	1339	1376	POTENTIAL.
FT	DOMAIN	429	599	CYTOPLASMIC (POTENTIAL).
FT	CARBOHYD	31	31	IMPORTANT FOR THE NEUROVIRULENCE.
FT	CARBOHYD	60	60	POTENTIAL.
FT	CARBOHYD	134	134	POTENTIAL.
FT	CARBOHYD	192	192	POTENTIAL.
FT	CARBOHYD	357	357	POTENTIAL.
FT	CARBOHYD	435	435	POTENTIAL.
FT	CARBOHYD	582	582	POTENTIAL.
FT	CARBOHYD	677	677	POTENTIAL.
FT	CARBOHYD	709	709	POTENTIAL.
FT	CARBOHYD	717	717	POTENTIAL.
FT	CARBOHYD	740	740	POTENTIAL.
FT	CARBOHYD	789	789	POTENTIAL.
FT	CARBOHYD	806	806	POTENTIAL.
FT	CARBOHYD	945	945	POTENTIAL.
FT	CARBOHYD	1232	1232	POTENTIAL.
FT	CARBOHYD	1242	1242	POTENTIAL.
FT	CARBOHYD	1261	1261	POTENTIAL.
FT	CARBOHYD	1277	1277	POTENTIAL.
FT	CARBOHYD	1298	1298	POTENTIAL.
FT	CARBOHYD	1370	1370	POTENTIAL.

SEQUENCE 1376 AA: 151882 MW: 9626615 CN:

CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score	=	5	Optimized Score	=	5	Significance	=	0.00
Residue Identity	=	100%	Matches	=	5	Mismatches	=	0
Gaps	=	0	Conservative Substitutions	=			=	0

						X	X
						YLNAT	
MLFVFILLLLPSCLGYIGDFRCI@TVNYNGNNASAPSISTEAVDVSKGLGTYVVLDRVYLNATLLLLTGYYYPVD	10	20	30	40	50	60	70
						X	
GSNYRNALALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNTPGTGATSYFPTIVIGSLFGNTSYTVVLEPY	80	90	100	110	120	130	140
NNIIMASVCTYTICQLPY	150	160					

14. CELSA-14 (1-5)

VGIH59 E2 glycoprotein precursor - Murine hepatitis virus

ENTRY	VGIH59	#Type Protein
TITLE	E2 glycoprotein precursor - Murine hepatitis virus (strain A59)	
ALTERNATE-NAME	peplomer glycoprotein\ spike glycoprotein	
DATE	31-Mar-1989 #Sequence 31-Mar-1989 #Text 30-Sep-1992	
PLACEMENT	2231.0	4.0 1.0 1.0 1.0
SOURCE	murine hepatitis virus, MHV	
ACCESSION	A27402	
REFERENCE		

#Authors Luytjes W., Sturman L.S., Brederbeek H.J., Charite J., van der Zeijst R.A.M., Horvink M.C., Spaan W.J.N.

#Journal Virology (1987) 161:479-487

#Title Primary structure of the glycoprotein E2 of coronavirus MHV-29 and identification of the

Best Available Copy

Accession: A27402
 #Molecule-type genomic RNA
 #Residues 1-1324 <LUY>
 COMMENT This virus is a member of the family Coronaviridae.
 SUPERFAMILY #Name coronavirus E2 glycoprotein
 KEYWORDS glycoprotein
 FEATURE

1-16 #Domain signal sequence <SIG>\ .
 17-1324 #Protein E2 glycoprotein <E2G>\
 17-717 #Protein 90B glycoprotein <EGB>\
 718-1324 #Protein 90A glycoprotein <EGA>\
 31,60,192,247,357,435,
 442,530,625,657,665,
 688,737,754,844,893,
 1126,1180,1190,1209,
 1225,1246,1318 #Binding-site carbohydrate (Asn)
 (covalent) (predicted)\
 1266-1286 #Domain transmembrane <TMN>
 SUMMARY #Molecular-weight 145963 #Length 1324 #Checksum 7500
 SEQUENCE

COMMENT Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

MLFVFILFLPSCLG YIGDFRCIQLVNSNGANVSAPSISTETVEVSQGLGTYYYVLDRVYLNATLLLTGYYPVD
 10 20 30 40 50 60 X 70
 YLNAT
 I I I I I
 GSKFRNLALTGTNSVLSWFOPPYLNQFNDGIFAKVQNLKTSTPSGATAYFFTIVIGSLFGYTSYTVVIEPY
 80 90 100 110 120 130 140
 NGVIMASVCQYTICQLPY
 150 160

15. CELSA-14 (1-5)

VGL2_CVMA5 E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PE

ID VGL2_CVMA5 STANDARD; PRT: 1324 AA.
 AC P11224;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
 DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
 GN S.
 OS MURINE CORONAVIRUS MHV (STRAIN A59).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; CORONAVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 88072088
 RA LUYTJES W., STURMAN L.B., BREDENBEERK P.J., CHARITE J.,
 RA VAN DER ZEIJST B.A.M., HORZINEK M.C., SPAAN W.J.M.;
 RL VIROLOGY 161:479-487(1987).
 CC -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
 CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE GLYCOPROTEIN.
 DR EMBL: M18379; CORMHVE2.
 DR PIR: A27402; UG1H59.
 KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL
 FT SIGNAL 1 1A
 FT CHAIN 1 1324

FT DOMAIN 718 1454 SPIKE PROTEIN S2 (90A).
FT TRANSMEM 1266 1286 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1287 1324 CYTOPLASMIC (POTENTIAL).
FT CARBOHYD 31 31 POTENTIAL.
FT CARBOHYD 60 60 POTENTIAL.
FT CARBOHYD 192 192 POTENTIAL.
FT CARBOHYD 357 357 POTENTIAL.
FT CARBOHYD 435 435 POTENTIAL.
FT CARBOHYD 530 530 POTENTIAL.
FT CARBOHYD 625 625 POTENTIAL.
FT CARBOHYD 657 657 POTENTIAL.
FT CARBOHYD 665 665 POTENTIAL.
FT CARBOHYD 688 688 POTENTIAL.
FT CARBOHYD 737 737 POTENTIAL.
FT CARBOHYD 754 754 POTENTIAL.
FT CARBOHYD 893 893 POTENTIAL.
FT CARBOHYD 1180 1180 POTENTIAL.
FT CARBOHYD 1190 1190 POTENTIAL.
FT CARBOHYD 1209 1209 POTENTIAL.
FT CARBOHYD 1225 1225 POTENTIAL.
FT CARBOHYD 1246 1246 POTENTIAL.
FT CARBOHYD 1318 1318 POTENTIAL.
SQ SEQUENCE 1324 AA; 145964 MW; 9039956 CN;
CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:02-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

MLFVFILFLPSCLG YIGDFRCIQLVNSNGANVSAPSI STETVEVSQGLGTYYVLDRVYLNATLLLTGYYPVD
10 20 30 40 50 60 X 70
GSKFRNLALTGTNSVSLSWFQPPYLNQFNDGIFAKVQNLKTS TP SGATAYFPTIVIGSLFCYTSYTVVIEPY
80 90 100 110 120 130 140
NGVIMASVCQYTICQLPY
150 160

16. CELSA-14 (1-5)

VG1HMJ E2 glycoprotein precursor - Murine hepatitis virus

ENTRY VG1HMJ #Type Protein
TITLE E2 glycoprotein precursor - Murine hepatitis virus
(strain JHM)
ALTERNATE-NAME peplomer glycoprotein\ spike glycoprotein
INCLUDES 90B glycoprotein\ 90A glycoprotein
DATE 31-Mar-1991 #Sequence 31-Mar-1991 #Text 30-Sep-1992
PLACEMENT 2231.0 4.0 1.0 2.0 1.0
SOURCE murine hepatitis virus, MHV
ACCESSION A33095
REFERENCE
#Authors Schmidt I., Skinner M., Siddell S.
#Journal J. Gen. Virol. (1987) 68:47-55
#Title Nucleotide sequence of the gene encoding the glycoprotein
projection glycoprotein of coronavirus JHMV-JHM.
#Reference-number A33095
#Accession A33095
#Molecule-type genomic RNA
#Accession J01330 CDS
COMMENT

FEATURE

1-10 #Domain signal sequence (SIG)\
 11-1235 #Protein E2 glycoprotein (E2G)\
 11-628 #Protein 90B glycoprotein (EGB)\
 629-1235 #Protein 90A glycoprotein (EGA)\
 1175-1208 #Domain transmembrane (TMN)\
 31,60,134,192,357,435,
 442,536,568,576,599,
 648,665,755,804,1037,
 1091,1101,1120,1136,
 1157,1229

#Binding-site carbohydrate (Asn)
 (covalent) (predicted)

SUMMARY

#Molecular-weight 136653 #Length 1235 #Checksum 9154

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:29:03-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

MLFVFILLLPSCLDYIGDFRCIQTVNYNGNNASAPSISTEAVDVSKGRGTYYVLDRVYLNATLLLTGYYPVD
 10 20 30 40 50 60 X 70
 YLNAT
 I I I I I
 GSNYRNALALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNTPTCATSYFPTJIVIGSLFGNTSYTVVLEPY
 80 90 100 110 120 130 140
 NNIIMASVCTYTICQLPY
 150 160

17. CELSA-14 (1-5)

VGL2_CVMJH E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PE

ID VGL2_CVMJH STANDARD; PRT: 1235 AA.
 AC P11225;
 DT 01-JUL-1989 (REL. 11, CREATED)
 DT 01-JUL-1989 (REL. 11, LAST SEQUENCE UPDATE)
 DT 01-MAY-1992 (REL. 22, LAST ANNOTATION UPDATE)
 DE E2 GLYCOPROTEIN PRECURSOR (SPIKE GLYCOPROTEIN) (PEPLOMER PROTEIN).
 GN S.
 OS MURINE CORONAVIRUS MHV (STRAIN JHM).
 OC VIRIDAE; SS-RNA ENVELOPED VIRUSES; POSITIVE-STRAND; CORONAVIRIDAE.
 RN [1]
 RP SEQUENCE FROM N.A.
 RM 87111467
 RA SCHMIDT I., SKINNER M.A., SIDDELL S.G.:
 RL J. GEN. VIROL. 68:47-56(1987).
 CC -!- FUNCTION: THE PEPLOMER PROTEIN MEDIATES THE BINDING OF VIRIONS
 CC TO THE HOST CELL RECEPTOR AND IS INVOLVED IN MEMBRANE FUSION.
 CC -!- SUBCELLULAR LOCATION: PERIPHERAL MEMBRANE GLYCOPROTEIN.
 DR EMBL: X04797; COMHVS.
 DR PIR: A33095; VGIHMJ.
 KW GLYCOPROTEIN; ENVELOPE PROTEIN; TRANSMEMBRANE; SIGNAL.
 FT SIGNAL 1 10
 FT CHAIN 11 1235
 FT CHAIN 11 628
 FT CHAIN 629 1235
 FT DOMAIN 11 1174
 FT TRANSMEM 1175 1208
 FT DOMAIN 12 1235
 FT CARBOHYD 11 11

FT CARBOHYD 134 134 POTENTIAL.
 FT CARBOHYD 192 192 POTENTIAL.
 FT CARBOHYD 357 357 POTENTIAL.
 FT CARBOHYD 435 435 POTENTIAL.
 FT CARBOHYD 536 536 POTENTIAL.
 FT CARBOHYD 568 568 POTENTIAL.
 FT CARBOHYD 576 576 POTENTIAL.
 FT CARBOHYD 599 599 POTENTIAL.
 FT CARBOHYD 648 648 POTENTIAL.
 FT CARBOHYD 665 665 POTENTIAL.
 FT CARBOHYD 804 804 POTENTIAL.
 FT CARBOHYD 1091 1091 POTENTIAL.
 FT CARBOHYD 1101 1101 POTENTIAL.
 FT CARBOHYD 1120 1120 POTENTIAL.
 FT CARBOHYD 1136 1136 POTENTIAL.
 FT CARBOHYD 1157 1157 POTENTIAL.
 FT CARBOHYD 1229 1229 POTENTIAL.
 SQ SEQUENCE 1235 AA: 136653 MW: 7753984 CN:
 CC -!- Retrieved by alexk on Thu 25 Feb 93 10:29:03-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

MLFVFILLLLPSCLGYIGDFRCIQTNYNGNNASAPSISTEAVDVSKGRGTYVVLDRVYLNATLLLTGYYPVD
 10 20 30 40 50 60 X 70
 YLNAT
 GSNYRNALALTGTNTLSLTWFKPPFLSEFNDGIFAKVQNLKTNTPTCATSYFPTIVIGSLFGNTSYTVVLEPY
 80 90 100 110 120 130 140
 NNIIMASVCTYTICQLPY
 150 160

18. CELSA-14 (1-5)

A40986 #M-cadherin - Mouse (fragment)

ENTRY A40986 #Type Protein (fragment)
 TITLE #M-cadherin - Mouse (fragment)
 DATE 16-Apr-1992 #Sequence 16-Apr-1992 #Text 16-Apr-1992
 PLACEMENT 0.0 0.0 0.0 0.0 0.0
 COMMENT #This entry is not verified.
 SOURCE Mus musculus #Common-name house mouse
 REFERENCE

#Authors Donalies M., Craner M., Ringwald M.,
 Stanzinski-Powitz A.
 #Journal Proc. Natl. Acad. Sci. U.S.A. (1991) 88:8024-8028
 #Title Expression of M-cadherin, a member of the cadherin
 multigene family, correlates with differentiation
 of skeletal muscle cells.

#Reference-number A40986
 #Accession A40986
 #Cross-reference GRIN74541

SUMMARY

SEQUENCE

COMMENT

Length 230 Amino acids 4198

Retrieved by alexk on Thu 25 Feb 93 10:29:03-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
 Residue Identity = 100% Matches = 5 Mismatches = 0
 Gaps = 0 Conservative Substitutions = 0

ENHKRLPYPLVQIKSDKQQLGSGVIYSIGPGVDDEPRNVFSIDKFTGRVYVLNATLDREKIDRFRLRAFALDL
 10 20 30 40 50 X 60 70
 GGSTLEDPTDLEIVVVDQNDNRPAFLQDVFRCHILEGAIPTGTFVTRAEATDADDPETDNAALRFSILEQGSP
 80 90 100 110 120 130 140

EFFSIDHTG
 150

19. CELSA-14 (1-5)

SAHU4F Cell surface antigen 4F2 heavy chain - Human

ENTRY SAHU4F #Type Protein
 TITLE Cell surface antigen 4F2 heavy chain - Human
 DATE 30-Jun-1990 #Sequence 30-Jun-1990 #Text 30-Jun-1992
 PLACEMENT 962.0 1.0 1.0 1.0 1.0
 SOURCE Homo sapiens #Common-name man
 ACCESSION A28455\ A28314\ A30240
 REFERENCE

#Authors Teixeira S., Di Grandi S., Kuehn L.C.
 #Journal J. Biol. Chem. (1987) 262:9574-9580
 #Title Primary structure of the human 4F2 antigen heavy chain predicts a transmembrane protein with a cytoplasmic NH2 terminus.

#Reference-number A28455

#Accession A28455

#Molecule-type mRNA

#Residues 1-529 <TEI>

REFERENCE

#Authors Quackenbush E., Clabby M., Gottesdiener K.M.,
 Barbosa J., Jones N.H., Strominger J.L., Speck S.,
 Leiden J.M.

#Journal Proc. Natl. Acad. Sci. U.S.A. (1987) 84:6526-6530

#Title Molecular cloning of complementary DNAs encoding the heavy chain of the human 4F2 cell-surface antigen: a type II membrane glycoprotein involved in normal and neoplastic cell growth.

#Reference-number A28314

#Accession A28314

#Molecule-type mRNA

#Residues 1-529 <QUA>

#Comment The authors' translation, given in Fig. 5, is missing residues 92-94, while the nucleotide sequence is missing the codons for residues 121-123. Consequently, the codons for residues 92-120 are associated with the translation corresponding to residues 95-123.

REFERENCE

#Authors Gottesdiener K.M., Karpinski B.A., Lindsten T.,
 Strominger J.L., Jones N.H., Thompson C.B., Leiden J.M.

#Journal Mol. Cell. Biol. (1988) 8:3809-3819

#Title Isolation and structural characterization of the human 4F2 heavy-chain gene, an inducible gene involved in T-lymphocyte activation.

#Reference-number A30240

#Accession A30240

#Molecule-type DNA

#Residues 1-529 <GGT>

#Comment The nucleotide sequence in Fig. 1 is inconsistent with the authors' translation in having the codon TTT (Phe) for Ser (219) and GAT (Glu) for Glu (441).

proliferation and transformed cells. It may be involved in the regulation of intracellular calcium concentration and the concomitant control of growth, excitability, and endocrine secretion. The molecule is a heterodimer consisting of a glycosylated heavy chain and a nonglycosylated light chain linked by at least one disulfide bond.

COMMENT

GENETIC

#Map-position 11q
#Name MDU1
#Introns 142/1, 200/1, 230/3, 253/3, 273/2, 333/3, 381/3, 409/3

SUPERFAMILY

KEYWORDS

#Name cell surface antigen 4F2 heavy chain
glycoprotein\ heterodimer\ surface antigen\
transmembrane protein

FEATURE

1-81 #Domain cytoplasmic <CYT>\
82-104 #Domain transmembrane <TMN>\
105-529 #Domain extracellular <EXT>\
264,280,323,405 #Binding-site carbohydrate (Asn)
(covalent) (predicted)

SUMMARY

#Molecular-weight 57944 #Length 529 #Checksum 8988

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 93 10:29:03-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

QVDTVATKVKDALEFWLQAGVDGFGVRDIENLKDASSFLAEWQNIKGFSEDRLLIAGTNSDDLQOILSLLE
230 240 250 260 270 280 290

X X

YLNAT

|||||

SNKDLLLTSSYLSDSGSTGEHTKSLVTQYLNATGNRWCSWSLSQARLLTSFLPAQLRLRYQLMLFTLPQTPV
300 310 320 X 330 340 350 360

FSYGDEIGLDAAALPGQPNAPVNLWDESSFPDIPGAVSANMTVKGOSEDPCSLLSLFRRL
370 380 390 400 410 420

20. CELSA-14 (1-5)

A39984

*Cell surface antigen 4F2 heavy chain - Human

ENTRY A39984 #Type Protein
TITLE *Cell surface antigen 4F2 heavy chain - Human
DATE 24-Jan-1992 #Sequence 24-Jan-1992 #Text 24-Jan-1992
PLACEMENT 0.0 0.0 0.0 0.0 0.0
COMMENT *This entry is not verified.
SOURCE Homo sapiens #Common-name nan
REFERENCE

#Authors Quackenbush E., Clabby M., Gutterdiere, F.M.,
Barbosa J., Jones N.H., Strominger, J.L., Clark D.,
Leiden J.M.

#Journal Proc. Natl. Acad. Sci. U.S.A. 11992 89:3436-3440

#Reference-number A39984

#Accession A39984

SUMMARY

#Molecular-weight 57944 #Length 529 #Checksum 8988

SEQUENCE

COMMENT

Retrieved by alexk on Thu 25 Feb 92 10:29:03-PST using FastDB

Initial Score = 5 Optimized Score = 5 Significance = 0.00
Residue Identity = 100% Matches = 5 Mismatches = 0
Gaps = 0 Conservative Substitutions = 0

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ **BLACK BORDERS**
- ☐ **IMAGE CUT OFF AT TOP, BOTTOM OR SIDES**
- ☒ **FADED TEXT OR DRAWING**
- ☐ **BLURRED OR ILLEGIBLE TEXT OR DRAWING**
- ☐ **SKEWED/SLANTED IMAGES**
- ☐ **COLOR OR BLACK AND WHITE PHOTOGRAPHS**
- ☐ **GRAY SCALE DOCUMENTS**
- ☐ **LINES OR MARKS ON ORIGINAL DOCUMENT**
- ☐ **REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY**
- ☐ **OTHER:** _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.

QVDTVATKVKDALEFWLDAGVGGFQVRDIENLKDASSPLAEWONI TRGFSEDL LTAGHSSOL 29115015
230 240 250 260 270 280 290

X X

YLNAT

||||

SNKDULLTSSYLSDBSCSTGEHTKSLVTQYLNATGNRWCSWSLSQARLLTSFLPAQLLRLYQLMLFTLPGETPV
300 310 320 X 330 340 350 360

FSYGDEIGLDAAALPGOPMEAFVMLWDESSFPDIPCAVSANNTVKG05EDP05LLSLFRRL
370 380 390 400 410 420

Best Available Copy

SCORE	0	1	1	1	2	3	3	4	5	5	6
STDEV	0	1	1	1	2	3	3	4	5	5	6

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	6
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	3	1.08

Times:	CPU	Total Elapsed
	00:03:20.10	00:07:11.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4353

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

*See next
printout for
alignments*

4 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. POLG_TBEVW	GENOME POLYPROTEIN (CONTAINS:	3414	6	6	4.63	0
2. GNWVNE	Genome polyprotein - Tick-bor	3414	6	6	4.63	0
3. GNWVTB	Genome polyprotein - Tick-bor	3412	6	6	4.63	0
4. POLG_TBEVS	GENOME POLYPROTEIN (CONTAINS:	3412	6	6	4.63	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
*** 3 standard deviations above mean score ***						
5. CBPC_HUMAN	MAST CELL CARBOXYPEPTIDASE A	417	5	5	3.70	0
6. A39246	*Carboxypeptidase A precursor	417	5	5	3.70	0
7. S94912	*egpA protein - Yersinia ente	415	5	5	3.70	0
8. PT25_HACSU	PHOSPHOTRANSFERASE CATALYNE 11	430	5	5	3.70	0
9. A37933	*Phosphotransferase system	410	5	5	3.70	0
10. RPR_HMW11	PROBABLE CARBOXYPEPTIDASE A	410	5	5	3.70	0
11. D39028	GENOME POLYPROTEIN (CONTAINS:	3412	5	5	3.70	0

> 0 <
01 10 IntelliGenetics
> 0 <

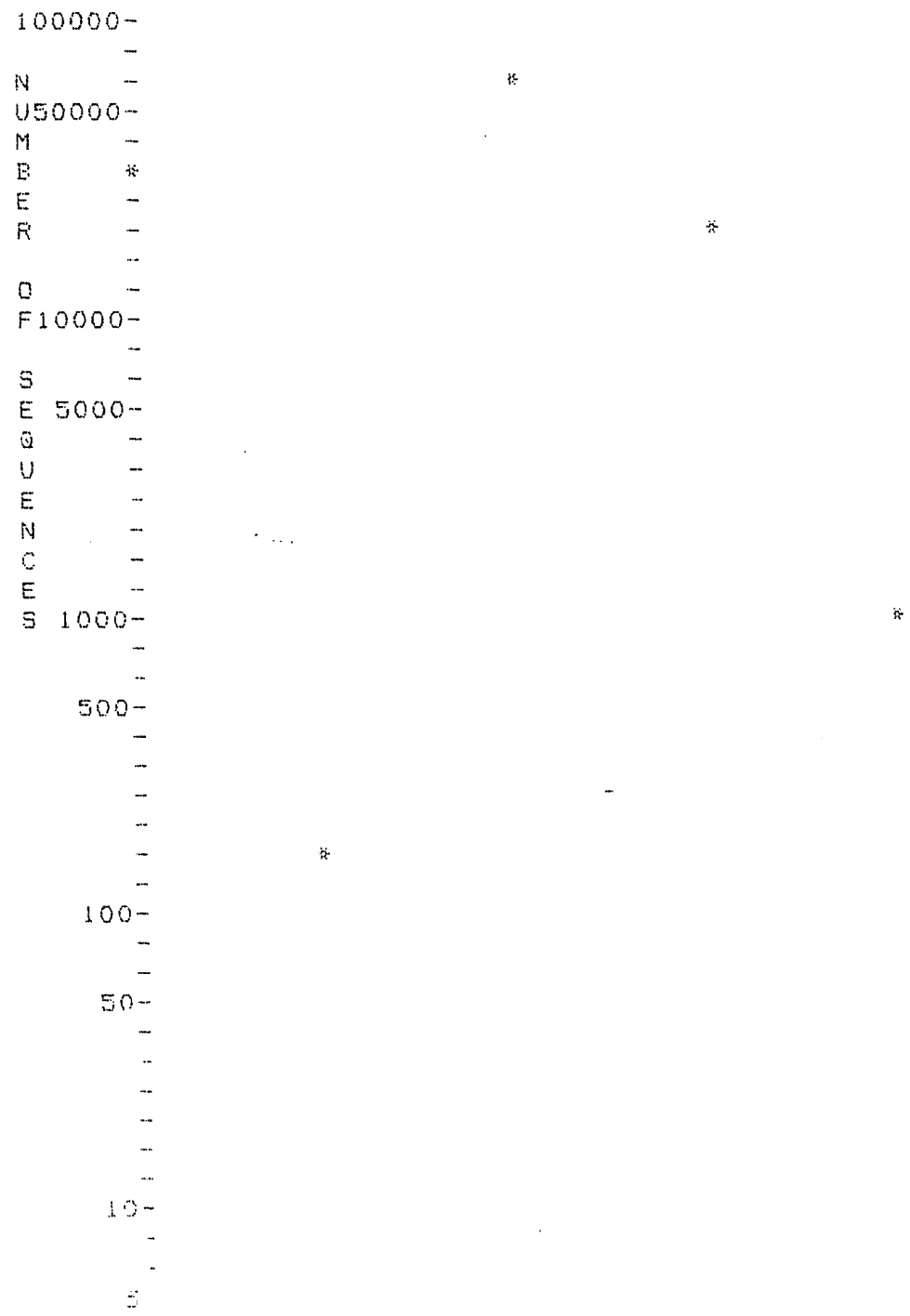
*Sequences numbered
as in table 4 1-19
(not by seq ID #)*

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-1.res made by alexk on Thu 25 Feb 93 10:21:13-PST.

Query sequence being compared: CELSA-1 (1-6)
Number of sequences searched: 94553
Number of scores above cutoff: 4353

Results of the initial comparison of CELSA-1 (1-6) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



14. DNBI_HSV1F	MAJOR DNA-BINDING PROTEIN.	1178	5	5	3.70	0
15. DNBI_HSV11	MAJOR DNA-BINDING PROTEIN.	1196	5	5	3.70	0
16. B30085	*Gene UL29 protein (major DNA	1196	5	5	3.70	0
17. DNBEHF	DNA-binding protein - Human h	1196	5	5	3.70	0
18. DNBEKS	DNA-binding protein - Human h	1196	5	5	3.70	0
19. DNBEV1	DNA-binding protein - Human h	1196	5	5	3.70	0
20. S19926	*Hypothetical protein - Red c	1864	5	5	3.70	0
**** 2 standard deviations above mean ****						
21. A36696	*Tumor necrosis factor - Mous	20	4	4	2.78	0
22. P50250	N-terminal of cystotatic prot	27	4	4	2.78	0
23. P98441	Sequence of C. trachomatis se	31	4	4	2.78	0
24. VG5_BPPH2	PROBABLE EARLY PROTEIN GP5.	66	4	4	2.78	0
25. ERBP59	Early protein gp5 (version 1)	66	4	4	2.78	0
26. A41608	*Hypothetical protein 1 - Rho	68	4	4	2.78	0
27. C552_BRAJA	CYTOCHROME C552.	73	4	4	2.78	0
28. PRT2_SEPOF	SPERMATID-SPECIFIC PROTEIN T2	77	4	4	2.78	0
29. B40973	*Spermatid-specific protein T	77	4	4	2.78	0
30. A27120	Epidermal growth factor-bind	77	4	4	2.78	0
31. PRT1_SEPOF	SPERMATID-SPECIFIC PROTEIN T1	78	4	4	2.78	0
32. A40973	*Spermatid-specific protein T	78	4	4	2.78	0
33. HPIS_THIRD	HIGH POTENTIAL IRON-SULFUR PR	85	4	4	2.78	0
34. IHTFER	High potential iron-sulfur pr	85	4	4	2.78	0
35. P60677	Human seminal alpha-inhibin-5	92	4	4	2.78	0
36. S11892	*Hypothetical protein - Fruit	95	4	4	2.78	0
37. S18821	*Cytochrome c552 - Bradyrhizo	104	4	4	2.78	0
38. AFL1_PETMA	BLOOD PLASMA APOLIPOPROTEIN L	105	4	4	2.78	0
39. A26602	Apolipoprotein 1 precursor -	105	4	4	2.78	0
40. CUP4_DROME	LARVAL CUTICLE PROTEIN IV PRE	112	4	4	2.78	0
41. CUP3_DROME	LARVAL CUTICLE PROTEIN III PR	112	4	4	2.78	0
42. D25299	Cuticle protein IV precursor	112	4	4	2.78	0
43. C25299	Cuticle protein III precursor	112	4	4	2.78	0
44. R10046	Region of IBDV strain E contg	113	4	4	2.78	0
45. MOT1_HUMAN	MOTILIN PRECURSOR.	115	4	4	2.78	0
46. A33323	*Promotilin precursor - Human	115	4	4	2.78	0
47. S04489	*Motilin - Human	115	4	4	2.78	0
48. S00156	Motilin precursor - Human	115	4	4	2.78	0
49. P80459	Sequence corresp. to the nucl	115	4	4	2.78	0
50. PA2_NAJPA	PHOSPHOLIPASE A2 (EC 3.1.1.4)	118	4	4	2.78	0

No alignments saved.

> 0 <
0| 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-2.res made by alexk on Thu 25 Feb 93 10:37:07-PST.

Query sequence being compared: CELSA-2 (1-7)
Number of sequences searched: 94553
Number of scores above cutoff: 4289

Results of the initial comparison of CELSA-2 (1-7) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries

100000-
-
N - *
U50000-
M -
B -
E -
R *
-
D -
F10000- * *
-
S -
E 5000-
Q - *
U -
E -
N -
C -
E -
S 1000- *
-
-
500-
-
-
-
-
-
100- *
-
-
50-
-
-
-
-
10-
-
5-

Best Available Copy

```

0-----
SCORE 0| 1| 1| 1| 2| 3| 3| 4| 5| 5| 6
STDEV 0 1 1 2 3 3 4 5 5 6

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5      Joining penalty 20
Gap penalty           1.00      Window size   6
Gap size penalty      0.26
Cutoff score          0
Randomization group   0

Initial scores to save 50      Alignments to save 0
Optimized scores to save 0      Display context    0

```

SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              1          3          1.06

```

```

Times:      CPU      Total Elapsed
            00:03:25.08      00:07:03.00

```

```

Number of residues:      25433612
Number of sequences searched: 94553
Number of scores above cutoff: 4289

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 4 standard deviations above mean ****						
1. YXYB_CALSA	HYPOTHETICAL 10.7 KD PROTEIN	97	6	6	4.73	0
2. C37202	*Hypothetical protein, XynC 3	97	6	6	4.73	0
3. R05567	Plant elongation factor-1 alp	293	6	6	4.73	0
4. R05569	Plant elongation factor-1 alp	294	6	6	4.73	0
5. ADT1_YEAST	ADP.ATP CARRIER PROTEIN 1 (AD	309	6	6	4.73	0
6. A24849	ADP.ATP carrier protein AAC1	309	6	6	4.73	0
7. S13805	*Elongation factor eEF-1 alph	323	6	6	4.73	0
8. A37159	*Centrosphere protein, elonga	325	6	6	4.73	0
9. A29946	Elongation factor 1 alpha cha	327	6	6	4.73	0
10. EF1A_RHYAM	ELONGATION FACTOR 1-ALPHA (EF	412	6	6	4.73	0
11. EF1A_EUGGR	ELONGATION FACTOR 1-ALPHA (EF	445	6	6	4.73	0
12. S07724	*Elongation factor eEF-1 alph	445	6	6	4.73	0
13. EF1A_STYLE	ELONGATION FACTOR 1-ALPHA (EF	445	6	6	4.73	0
14. S16308	*Elongation factor eEF-1 alph	447	6	6	4.73	0

17. JS0719	Elongation factor 1 alpha cha	447	6	6	4.73	0
18. EF1A_LYCES	ELONGATION FACTOR 1-ALPHA (EF	448	6	6	4.73	0
19. S10507	*Elongation factor eEF-1 alph	448	6	6	4.73	0
20. S09355	*Elongation factor eEF-1 alph	448	6	6	4.73	0
21. R05565	Plant elongation factor-1 alp	448	6	6	4.73	0
22. EF1A_ARATH	ELONGATION FACTOR 1-ALPHA (EF	449	6	6	4.73	0
23. S21989	*Elongation factor 1A - Carro	449	6	6	4.73	0
24. S08348	*Elongation factor eEF-1 alph	449	6	6	4.73	0
25. S08534	*Elongation factor eEF-1 alph	449	6	6	4.73	0
26. A32684	Elongation factor 1 alpha cha	449	6	6	4.73	0
27. S06724	Elongation factor 1 alpha cha	449	6	6	4.73	0
28. S10226	*Elongation factor eEF-1 alph	454	6	6	4.73	0
29. EF1A_DICDI	ELONGATION FACTOR 1-ALPHA (EF	456	6	6	4.73	0
30. S11665	*Elongation factor eEF-1 alph	456	6	6	4.73	0
31. EF13_RHIRA	ELONGATION FACTOR 1-ALPHA (EF	457	6	6	4.73	0
32. EF1A_YEAST	ELONGATION FACTOR 1-ALPHA (EF	458	6	6	4.73	0
33. EF1A_CANAL	ELONGATION FACTOR 1-ALPHA (EF	458	6	6	4.73	0
34. EF12_RHIRA	ELONGATION FACTOR 1-ALPHA (EF	458	6	6	4.73	0
35. EF11_RHIRA	ELONGATION FACTOR 1-ALPHA (EF	458	6	6	4.73	0
36. A35154	*Elongation factor 1 alpha -	458	6	6	4.73	0
37. A25938	Elongation factor 1 alpha cha	458	6	6	4.73	0
38. S08058	Elongation factor 1 alpha-3 c	458	6	6	4.73	0
39. S06300	Elongation factor 1 alpha-2 c	458	6	6	4.73	0
40. EFBY1A	Elongation factor eEF-1 alpha	458	6	6	4.73	0
41. EF1A_MOUSE	ELONGATION FACTOR 1-ALPHA (EF	461	6	6	4.73	0
42. EF1A_ARTSA	ELONGATION FACTOR 1-ALPHA (EF	461	6	6	4.73	0
43. EF1A_APIME	ELONGATION FACTOR 1-ALPHA (EF	461	6	6	4.73	0
44. EF13_XENLA	ELONGATION FACTOR 1-ALPHA, 00	461	6	6	4.73	0
45. EF12_XENLA	ELONGATION FACTOR 1-ALPHA, 00	461	6	6	4.73	0
46. S13806	*Elongation factor eEF-1 alph	461	6	6	4.73	0
47. S10225	*Elongation factor eEF-1 alph	461	6	6	4.73	0
48. JH0530	Elongation factor 1-alpha 0 -	461	6	6	4.73	0
49. EFHB1	Elongation factor eEF-1 alpha	461	6	6	4.73	0
50. EFMS1	Elongation factor eEF-1 alpha	461	6	6	4.73	0

No alignments saved.

> 0 <
01 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-3.res made by alexk on Thu 25 Feb 93 10:28:59-PST.

Query sequence being compared: CELSA-3 (1-5)
Number of sequences searched: 94553
Number of scores above cutoff: 4004

Results of the initial comparison of CELSA-3 (1-5) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries

100000-

-

N -

* .

U50000-

M -

B *

E -

R -

-

D -

F10000-

*

-

S -

E 5000-

Q -

U -

E -

N -

C -

E -

S 1000-

-

-

500-

*

-

-

-

-

-

-

100-

-

-

50-

-

-

-

-

-

10-

-

-

5-

-

*

Best Available Copy

```

-
-
-
0-----
| | | | | | | | | | |
| | | | | | | | | | |
SCORE 0 | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5
STDEV 0 | 1 | 1 | 2 | 2 | 3 | 3 | 3 | 4 | 4 | 5

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5             Joining penalty 20
Gap penalty            1.00          Window size    5
Gap size penalty      0.26
Cutoff score          0
Randomization group   0

Initial scores to save 50      Alignments to save 0
Optimized scores to save 0    Display context    0

```

SEARCH STATISTICS

```

Scores:                Mean      Median      Standard Deviation
                        1          3          1.03

```

```

Times:                CPU          Total Elapsed
                   00:03:18.96    00:06:54.00

```

```

Number of residues:      25433612
Number of sequences searched: 94553
Number of scores above cutoff: 4004

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

5 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. PULA_KLEAE	PULLULANASE (EC 3.2.1.41) (AL	1096	5	5	3.89	0
2. A26879	alpha-Dextrin endo-1,6-alpha-	1096	5	5	3.89	0
3. P82507	Pullulanase protein.	1096	5	5	3.89	0
4. DBH_THETH	DNA-BINDING PROTEIN II.	95	5	5	3.89	0
5. S12888	*DNA-binding protein II - The	95	5	5	3.89	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 2 standard deviations above mean ****						
6. S10561	*Photosystem II 21K protein -	56	4	4	2.92	0
7. C551_PSEST	CYTOCHROME C551.	82	4	4	2.92	0
8. CCFB56	Cytochrome c551 - Pseudomonas	82	4	4	2.92	0
9. CYC6_PLEBO	CYTOCHROME C6 (SOLUBLE CYTOCH	85	4	4	2.92	0
10. CCFB5	Cytochrome c6 - Plectonera bo	85	4	4	2.92	0
11. CYC6_AHAYA	CYTOCHROME C6 (SOLUBLE CYTOCH	85	4	4	2.92	0

14. DNBS2F	DNA-binding protein II - Bac	92	4	4	2.92	0
15. DBH_BACSV	DNA-BINDING PROTEIN II (HB) (92	4	4	2.92	0
16. B39409	*DNA-binding protein HB - Bac	92	4	4	2.92	0
17. S00015	DNA-binding protein HB - Baci	104	4	4	2.92	0
18. R05446	CAT-GLP-1 hybrid protein.	105	4	4	2.92	0
19. PINO_ECOLI	PINO PROTEIN.	105	4	4	2.92	0
20. GLN1_METTL	NITROGEN FIXATION NIFHD REGIO	105	4	4	2.92	0
21. S08095	*Hypothetical protein (nifH1	105	4	4	2.92	0
22. S06985	Hypothetical protein (nifH1 3	105	4	4	2.92	0
23. Q0ECRP	Hypothetical protein A-105 -	110	4	4	2.92	0
24. YP12_RTBPV	HYPOTHETICAL P12 PROTEIN (ORF	110	4	4	2.92	0
25. YP12_RTBV	HYPOTHETICAL P12 PROTEIN (ORF	110	4	4	2.92	0
26. B40785	*ORF2 protein - Rice tungro b	110	4	4	2.92	0
27. S16667	*Hypothetical protein P12 - R	110	4	4	2.92	0
28. P70092	Sequence encoded by (2'-5') o	111	4	4	2.92	0
29. S18858	*Cytochrome c-553 - Anabaena	111	4	4	2.92	0
30. S22481	*Cytochrome c553 precursor -	111	4	4	2.92	0
31. P71705	Partial (2'-5') oligo A synth	112	4	4	2.92	0
32. N075_PEA	EARLY NODULIN-75 PROTEIN (N-7	112	4	4	2.92	0
33. S10101	*Nodulation protein ENOD2 - G	114	4	4	2.92	0
34. YCW8_YEAST	HYPOTHETICAL 12.4 KD PROTEIN	114	4	4	2.92	0
35. S19492	Hypothetical protein YCR78C -	132	4	4	2.92	0
36. R05445	CAT-A4-751i hybrid protein.	137	4	4	2.92	0
37. CY2_RHOCA	CYTOCHROME C2 PRECURSOR.	137	4	4	2.92	0
38. CCRF2C	Cytochrome c2 precursor - Rho	146	4	4	2.92	0
39. YZP2_ECOLI	VERY HYPOTHETICAL 16.1 KD PRO	146	4	4	2.92	0
40. A35720	*Hypothetical protein 146 (ph	160	4	4	2.92	0
41. PETD_PROHO	CYTOCHROME B6-F COMPLEX SUBUN	160	4	4	2.92	0
42. S16853	*petD protein chain IV - Proc	175	4	4	2.92	0
43. MAT1_YEAST	MATING HORMONE ALPHA-1.	175	4	4	2.92	0
44. COAG_LIMPO	COAGULOGEN (CONTAINS: COAGUL1	175	4	4	2.92	0
45. S19453	Mating hormone alpha-1 - Yeas	175	4	4	2.92	0
46. S19397	Mating hormone alpha-1 - Yeas	175	4	4	2.92	0
47. WCHCA	Coagulogen - Atlantic horsesh	175	4	4	2.92	0
48. JFBYA1	Mating hormone alpha-1 - Yeas	179	4	4	2.92	0
49. SPC4_CANFA	MICROSOMAL SIGNAL PEPTIDASE 1	179	4	4	2.92	0
50. A35309	*Signal peptidase 18K chain -					

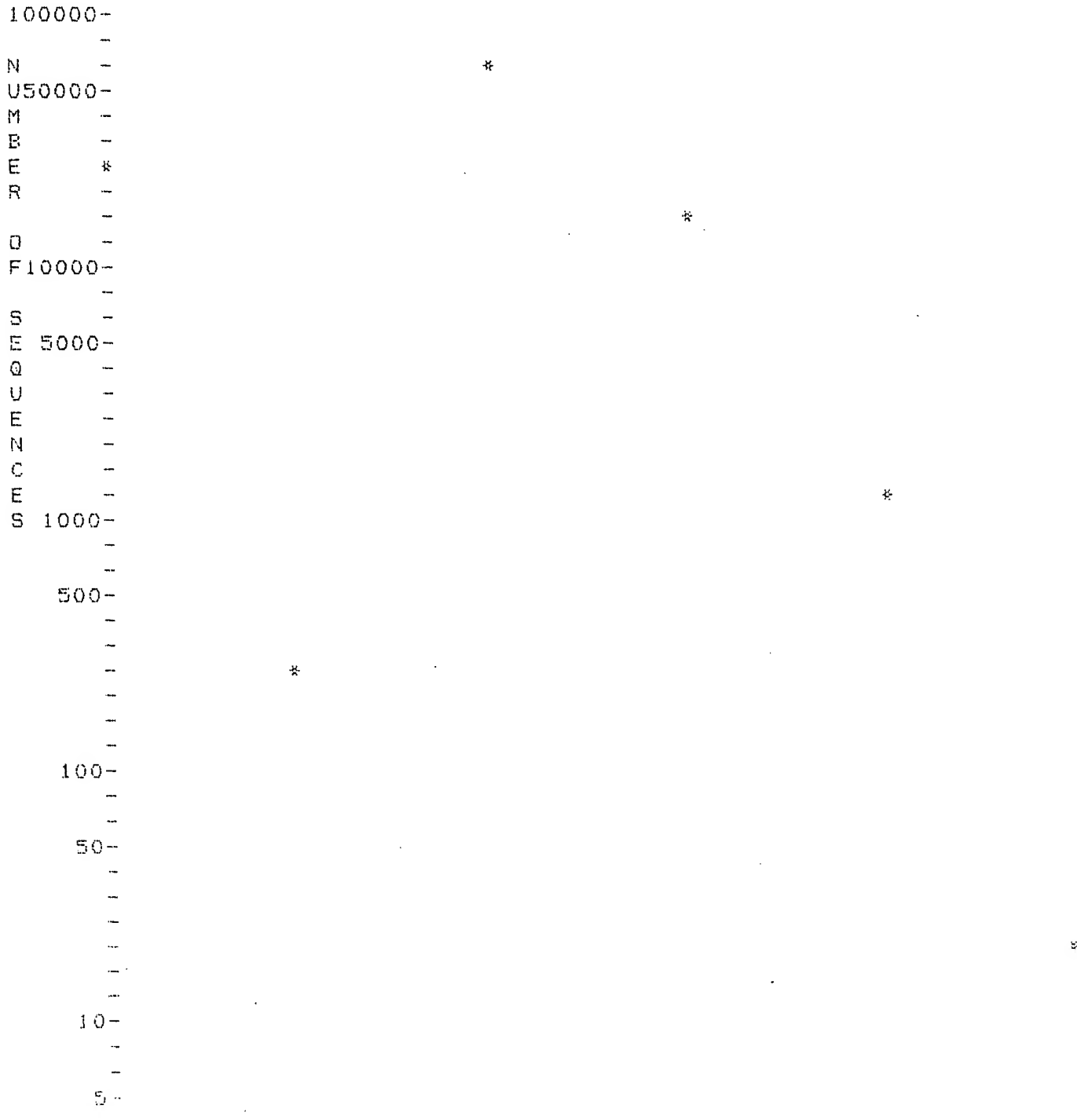
No alignments saved.

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-4.res made by alexk on Thu 25 Feb 93 10:41:56-PST.

Query sequence being compared: CELSA-4 (1-5)
Number of sequences searched: 94553
Number of scores above cutoff: 4622

Results of the initial comparison of CELSA-4 (1-5) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

```

-----
SCORE 0| 1| 1| 2| 2| 3| 3| 4| 4| 5
STDEV 0 1 1 2 3 3 4 4 5

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5      Joining penalty      20
Gap penalty      1.00      Window size      5
Gap size penalty      0.26
Cutoff score      0
Randomization group      0

Initial scores to save      50      Alignments to save      0
Optimized scores to save      0      Display context      0

```

SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              1          3          1.04

```

```

Times:      CPU      Total Elapsed
            00:03:29.98      00:03:54.00

```

```

Number of residues:      25433612
Number of sequences searched:      94553
Number of scores above cutoff:      4622

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.

```

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

```

17 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. S20590	*Sialidase - Actinomyces visc	913	5	5	3.84	0
2. A33827	*Regulatory protein ral2 - Ye	611	5	5	3.84	0
3. RAL2_SCHPO	RAL2 PROTEIN.	611	5	5	3.84	0
4. WMCVFM	Inclusion body matrix protein	512	5	5	3.84	0
5. IBMP_FMVD	INCLUSION BODY MATRIX PROTEIN	512	5	5	3.84	0
6. A30828	Steroid 17alpha-monooxygenase	507	5	5	3.84	0
7. S16719	*Steroid 17alpha-monooxygenas	507	5	5	3.84	0
8. CPT1_RAT	CYTOCHROME P450 XVIIIA1 (P450-	507	5	5	3.84	0
9. JH0594	Vasoactive intestinal polypep	459	5	5	3.84	0
10. S16562	*nOLF protein - Rhizobium mel	367	5	5	3.84	0
11. NOLF_RHIME	NODULATION PROTEIN NOLF.	367	5	5	3.84	0
12. A27659	Cytochrome P430 17 - Rat (Pna	337	5	5	3.84	0
13. A33980	Steroid 17alpha-monooxygenase	235	5	5	3.84	0
14. S18659	*Hypothetical protein - Mycop	143	5	5	3.84	0
15. R20793	CDR-grafted, humanised heavy	146	5	5	3.84	0
16. S06727	*Hypothetical protein 1 (mini	122	5	5	3.84	0
17. YN2_STRCD	MINI-CIRCLE HYPOTHETICAL 12.3	121	5	5	3.84	0

		Init. Opt.				
Sequence Name	Description	Length	Score	Score	Sig.	Frame

**** 2 standard deviations above mean ****						
18. A35776	*RecQ protein - Escherichia c	5	4	4	2.88	0
19. S04171	aadA protein - Klebsiella pne	23	4	4	2.88	0
20. RP30_YEAST	RIBOSOMAL PROTEIN RP30 (FRAGM	25	4	4	2.88	0
21. P90810	Synthetic HTLV-1 peptide anti	26	4	4	2.88	0
22. D41161	*31K antigen PEB4 - Campyloba	34	4	4	2.88	0
23. R13186	Peptide VII immunoreactive wi	38	4	4	2.88	0
24. R06408	HTLV-1 corresponding peptide	38	4	4	2.88	0
25. S10765	*Glutamate synthase - Azospir	40	4	4	2.88	0
26. R20772	Peptide V based on immunoreac	40	4	4	2.88	0
27. R13185	Peptide (V) immunoreactive wi	40	4	4	2.88	0
28. R13591	HTLV-I env precursor epitope	40	4	4	2.88	0
29. TYBB_SALGA	THYMOSIN BETA-12.	42	4	4	2.88	0
30. LCA_MACGI	ALPHA-LACTALBUMIN I (LACTOSE	42	4	4	2.88	0
31. B39880	*Streptomycin/spectinomycin r	42	4	4	2.88	0
32. A10265	alpha-Lactalbumin I - Eastern	42	4	4	2.88	0
33. THIO_EUBAC	THIOREDOXIN (FRAGMENT).	45	4	4	2.88	0
34. PSBK_CHLRE	PHOTOSYSTEM II 4 KD REACTION	46	4	4	2.88	0
35. S11162	Photosystem II protein psbK p	46	4	4	2.88	0
36. PSBK_MARPO	PHOTOSYSTEM II 4 KD REACTION	55	4	4	2.88	0
37. S01585	Photosystem II protein psbK p	55	4	4	2.88	0
38. A05024	Hypothetical protein 55 - Liv	55	4	4	2.88	0
39. PSBK_TOBAC	PHOTOSYSTEM II 4 KD REACTION	61	4	4	2.88	0
40. PSBK_SINAL	PHOTOSYSTEM II 4 KD REACTION	61	4	4	2.88	0
41. PSBK_DRYSA	PHOTOSYSTEM II 4 KD REACTION	61	4	4	2.88	0
42. PSBK_HORVU	PHOTOSYSTEM II 4 KD REACTION	61	4	4	2.88	0
43. S02115	Photosystem II protein psbK p	61	4	4	2.88	0
44. F2RZKS	Photosystem II protein psbK -	61	4	4	2.88	0
45. E26839	Streptomycin/spectinomycin re	63	4	4	2.88	0
46. R13345	P1684 HCV antigen (1684-1750)	67	4	4	2.88	0
47. A41474	*Heat-stable enterotoxin yst	71	4	4	2.88	0
48. S11707	*Aminoglycoside adenylyltrans	72	4	4	2.88	0
49. R10026	ompA Signal peptide used to c	78	4	4	2.88	0
50. CPR3_ECOLI	COPB PROTEIN (REPA2 PROTEIN).	84	4	4	2.88	0

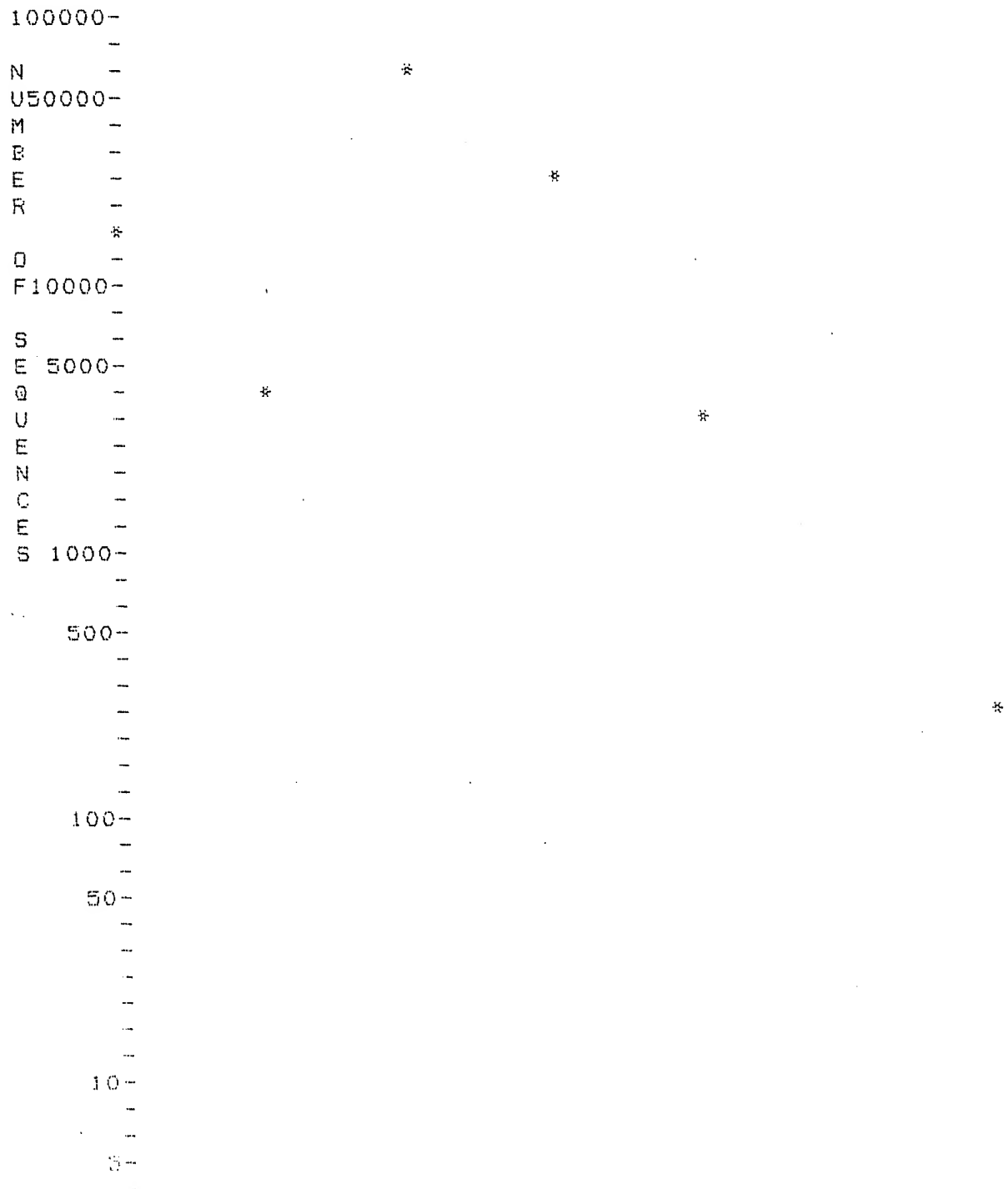
No alignments saved.

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-5.res made by alexk on Thu 25 Feb 93 10:21:15-PST.

Query sequence being compared: CELSA-5 (1-10)
Number of sequences searched: 94553
Number of scores above cutoff: 3851

Results of the initial comparison of CELSA-5 (1-10) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

SCORE	0	1	2	3	4	5	6	7
STDEV	0	1	2	3	4	5	6	7

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	3	1.01

Times:	CPU	Total Elapsed
	00:03:30.00	00:07:13.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	3851

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. LACP_KLULA	LACTOSE PERNEASE.	587	7	7	5.92	0
2. A31776	Lactose permease - Yeast (Klu	587	7	7	5.92	0
**** 4 standard deviations above mean ****						
3. S13898	*Alkaline phosphatase - Rabbi	16	6	6	4.93	0
4. MCP5_MOUSE	MAST CELL PROTEASE 5 (EC 3.4.	30	6	7	4.93	0
5. C35646	*Mast cell proteinase MMCP-5	30	6	7	4.93	0
6. R15700	Human tumour-associated cell	30	6	6	4.93	0
7. A28064	Proglucagon - Pig (fragment)	48	6	7	4.93	0
8. P91880	Fusion protein coded for by o	50	6	7	4.93	0
9. P92100	Fusion protein coded for by o	63	6	7	4.93	0
10. VES6_BPV4	PROBABLE E6 PROTEIN.	99	6	6	4.93	0
11. W6W1B4	E6 protein - Bovine papilloma	99	6	6	4.93	0
12. B40093	*Ubiquitin carboxyl-terminal	113	6	6	4.93	0
13. CXC_GAINT	CYTOKINE G	111	6	6	4.93	0

16. CYC_NIGDA	CYTOCHROME C.	111	6	6	4.93	0
17. CYC_MAIZE	CYTOCHROME C.	111	6	6	4.93	0
18. CYC_HELAN	CYTOCHROME C.	111	6	6	4.93	0
19. CYC_CUCMA	CYTOCHROME C.	111	6	6	4.93	0
20. CYC_BRAOL	CYTOCHROME C.	111	6	6	4.93	0
21. CYC_ARUMA	CYTOCHROME C.	111	6	6	4.93	0
22. CCFS	Cytochrome c - Common sunflow	111	6	6	4.93	0
23. CCND	Cytochrome c - Love-in-a-mist	111	6	6	4.93	0
24. CCRM	Cytochrome c - Cuckoopint	111	6	6	4.93	0
25. CCED	Cytochrome c - European elder	111	6	6	4.93	0
26. CCZM	Cytochrome c - Maize	111	6	6	4.93	0
27. CCRZ	Cytochrome c - Rice	111	6	6	4.93	0
28. CCMB	Cytochrome c - Mung bean	111	6	6	4.93	0
29. CCPU	Cytochrome c - Cucurbit	111	6	6	4.93	0
30. CCRPBO	Cytochrome c - Cauliflower	111	6	6	4.93	0
31. CCRPBN	Cytochrome c - Rape	111	6	6	4.93	0
32. CYC_WHEAT	CYTOCHROME C.	112	6	6	4.93	0
33. JS0709	Cytochrome c - Rice	112	6	6	4.93	0
34. CCWT	Cytochrome c - Wheat	112	6	6	4.93	0
35. S21626	*Collagen alpha 1(I) chain -	147	6	6	4.93	0
36. P70557	Tumor necrosis factor.	154	6	6	4.93	0
37. P60526	Sequence of tumour necrosis f	156	6	6	4.93	0
38. P60527	Sequence of tumour necrosis f	157	6	6	4.93	0
39. GLUC_FIG	GLUCAGON PRECURSOR (FRAGMENT)	158	6	7	4.93	0
40. P70558	Tumor necrosis factor.	170	6	6	4.93	0
41. YCED_ECOLI	HYPOTHETICAL 19.3 KD PROTEIN	173	6	6	4.93	0
42. JV0047	Hypothetical 30K protein (rpm	173	6	6	4.93	0
43. GLUC_OCTDE	GLUCAGON PRECURSOR.	180	6	7	4.93	0
44. GLUC_MESAU	GLUCAGON PRECURSOR.	180	6	7	4.93	0
45. C36118	*Proglucagon precursor - Degu	180	6	7	4.93	0
46. GCHY	Proglucagon precursor - Golde	180	6	7	4.93	0
47. R04190	Peptide derived from gp120 en	180	6	6	4.93	0
48. TRIF_CHICK	TROPONIN I, FAST SKELETAL MUS	182	6	6	4.93	0
49. TPCHIS	Troponin I, fast skeletal mus	182	6	6	4.93	0
50. A24918	Troponin I, fast skeletal mus	183	6	6	4.93	0

No alignments saved.

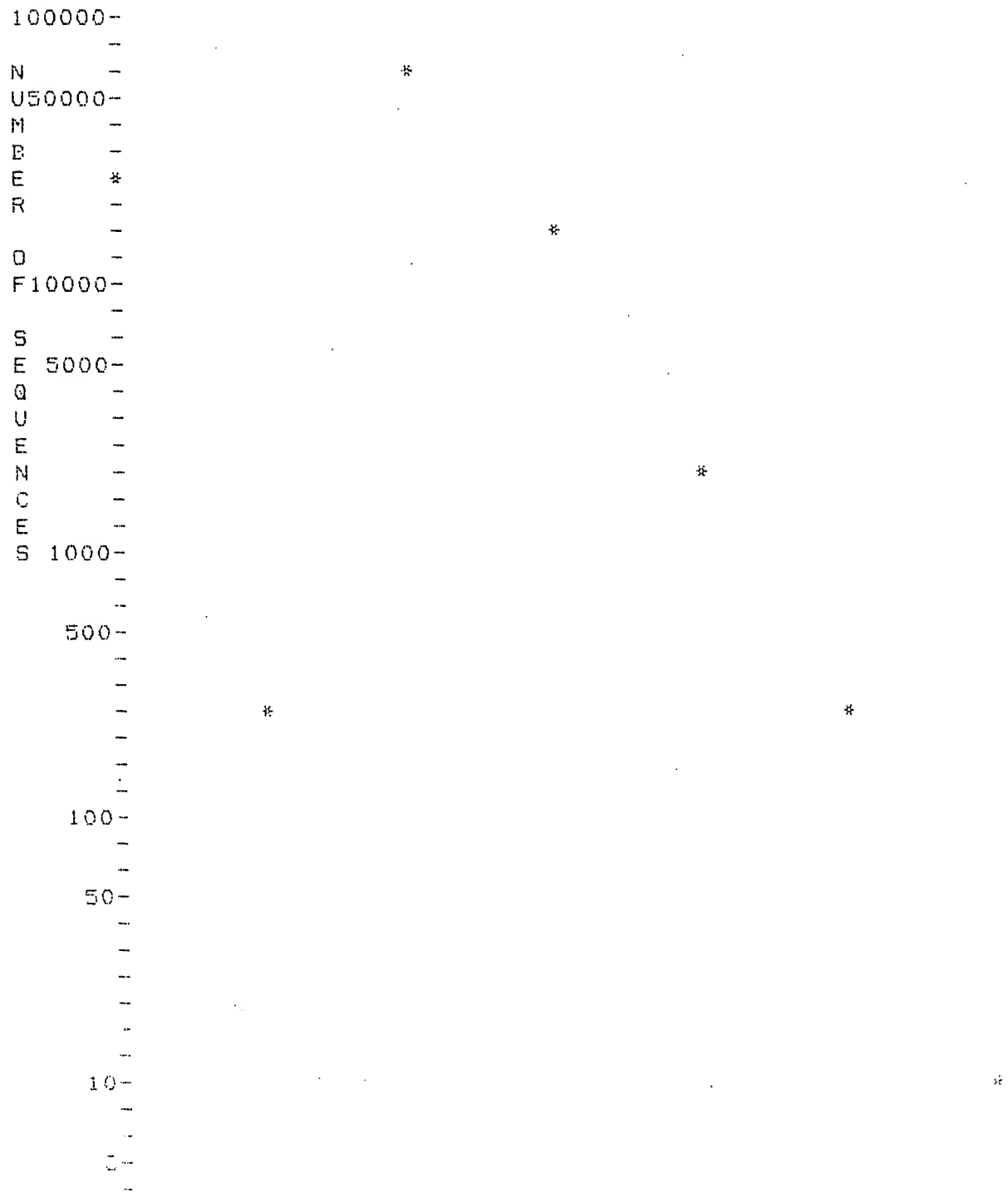
> 0 <
0| 0| IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-6.res made by alexk on Thu 25 Feb 93 10:28:59-PST.

Query sequence being compared: CELSA-6 (1-9)
Number of sequences searched: 94553
Number of scores above cutoff: 3802

Results of the initial comparison of CELSA-6 (1-9) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

	0	1	2	3	4	5	6	7
SCORE	0	1	2	3	4	5	6	7
STDEV	0	1	2	3	4	5	6	7

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	3	1.05

Times:	CPU	Total Elapsed
	00:03:24.97	00:06:53.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	3802

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. VG01_BPP22	PORTAL PROTEIN (PROTEIN GPI),	724	7	7	5.70	0
2. Z1BP22	Cene 1 protein - Phage P22	725	7	7	5.70	0
**** 4 standard deviations above mean ****						
3. RECO_BOVIN	RECOVERIN (P26),	202	6	6	4.75	0
4. S21155	*23K protein - Mouse	202	6	6	4.75	0
5. A38433	*Recoverin - Bovine	202	6	6	4.75	0
6. S19305	*Calcium binding protein P26	202	6	6	4.75	0
7. ATC2_YEAST	PROBABLE CALCIUM-TRANSPORTING	1091	6	6	4.75	0
8. PWBYSR	Ca2+-transporting ATPase PWB2	1091	6	6	4.75	0
9. CYGS_STRPU	SPERACT RECEPTOR PRECURSOR (C	1125	6	6	4.75	0
10. A33535	*Guanylate cyclase precursor	1125	6	6	4.75	0
11. A30856	*Guanylate cyclase - Sea urch	1125	6	6	4.75	0
**** 3 standard deviations above mean ****						
12. P591_SAT	PERLIN 1 (CUB-2) (FRAGMENT)	10	5	5	3.34	0

15. VKIL_BPP22	KIL PROTEIN.	82	5	5	3.80	0
16. VDBP22	kil protein - Phage P22	82	5	5	3.80	0
17. FER1_EOUTE	FERREDOXIN I.	95	5	5	3.80	0
18. FER1_EQUAR	FERREDOXIN I.	95	5	5	3.80	0
19. FEEQ1F	Ferredoxin I - Field horsetail	95	5	5	3.80	0
20. FEEQ1	Ferredoxin I - Horsetail (Equ	95	5	5	3.80	0
21. CYTC_BOVIN	CYSTATIN, COLOSTRUM (THIOL PR	112	5	6	3.80	0
22. UDBO	Cystatin - Bovine	112	5	6	3.80	0
23. P30180	Sequence of a polypeptide wit	155	5	5	3.80	0
24. P40124	Sequence encoded by the cDNA	155	5	5	3.80	0
25. LEPA_PSEFL	LEPA PROTEIN (FRAGMENT).	161	5	5	3.80	0
26. P60691	Sequence of the region coding	161	5	5	3.80	0
27. P80055	Sequence of human interferon	161	5	5	3.80	0
28. P50281	Protein sequence encoding mod	164	5	5	3.80	0
29. P30227	Sequence of hybrid interferon	165	5	5	3.80	0
30. P50207	Modified human interferon (IF	165	5	5	3.80	0
31. P50280	Protein sequence encoding mod	165	5	5	3.80	0
32. P50273	Protein sequence encoding mod	165	5	5	3.80	0
33. S04479	*Interferon beta-1 - Human	166	5	5	3.80	0
34. P30222	Sequence of hybrid interferon	166	5	5	3.80	0
35. P30219	Sequence of interferon (HuIFN	166	5	5	3.80	0
36. P40626	Sequence of interferon-beta-S	166	5	5	3.80	0
37. P50206	Modified human interferon (IF	166	5	5	3.80	0
38. P50197	Modified human interferon (IF	166	5	5	3.80	0
39. P50196	Modified human interferon (IF	166	5	5	3.80	0
40. P50195	Modified human interferon (IF	166	5	5	3.80	0
41. P50194	Modified human interferon (IF	166	5	5	3.80	0
42. P50193	Modified human interferon (IF	166	5	5	3.80	0
43. P50192	Modified human interferon (IF	166	5	5	3.80	0
44. P50279	Protein sequence encoding syn	166	5	5	3.80	0
45. P50278	Protein sequence encoding mod	166	5	5	3.80	0
46. P50277	Protein sequence encoding mod	166	5	5	3.80	0
47. P50276	Protein sequence encoding mod	166	5	5	3.80	0
48. P50275	Protein sequence encoding mod	166	5	5	3.80	0
49. P50274	Protein sequence encoding mod	166	5	5	3.80	0
50. P50272	Protein sequence encoding mod	166	5	5	3.80	0

No alignments saved.

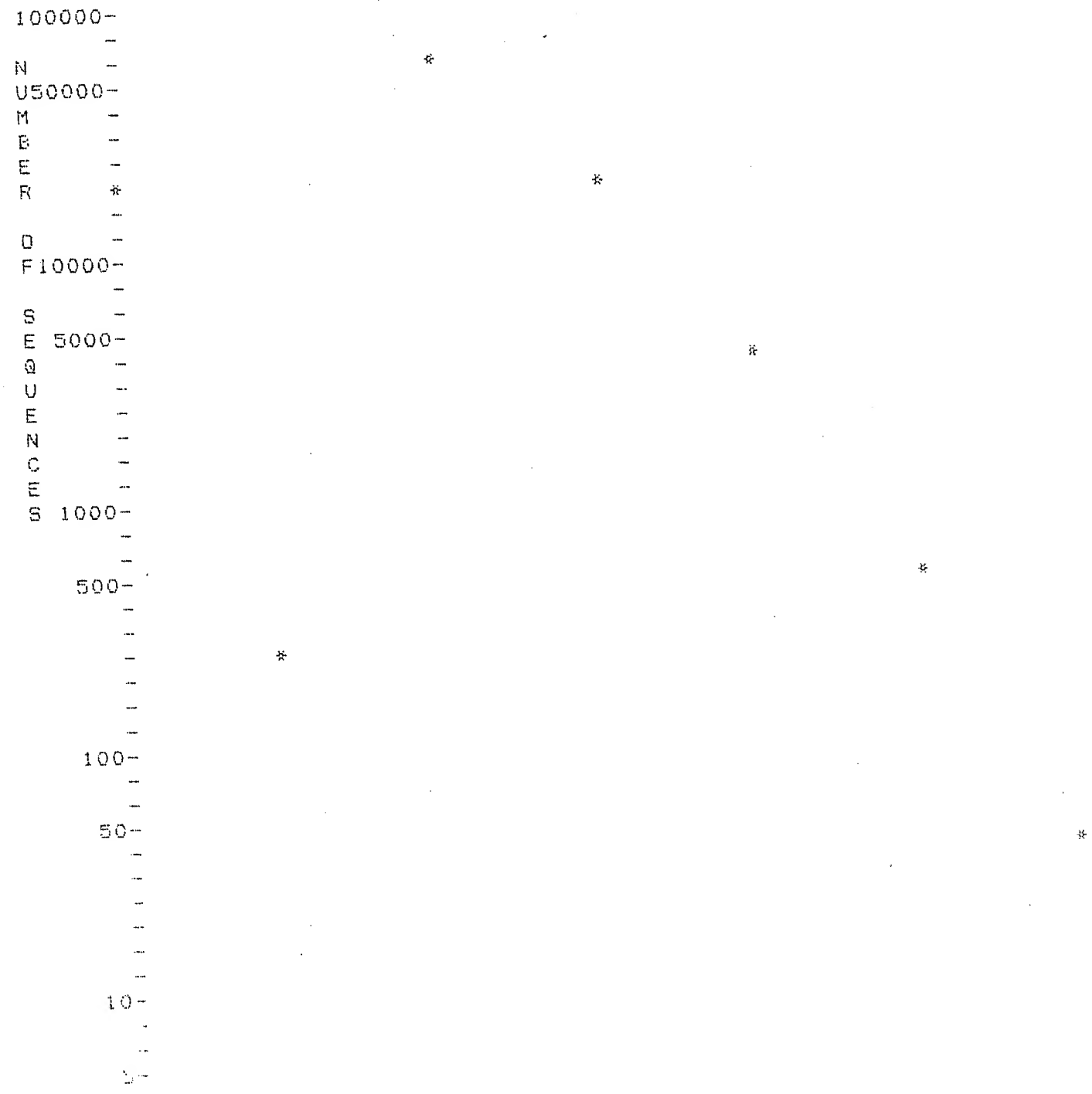
> 0 <
01 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-7.res made by alexk on Thu 25 Feb 93 10:37:07-PST.

Query sequence being compared: CELSA-7 (1-10)
Number of sequences searched: 94553
Number of scores above cutoff: 3985

Results of the initial comparison of CELSA-7 (1-10) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

```

0-----
||      |      |      |      |      |      |      |      |      |      |
SCORE 0|      1      |      1      |      2|      3      |      3      |      4|      5      |      5      |      6
STDEV  0      1      |      1      |      2      |      3      |      3      |      4      |      5      |      5      |      6

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5      Joining penalty      20
Gap penalty      1.00      Window size      5
Gap size penalty      0.26
Cutoff score      0
Randomization group      0

```

```

Initial scores to save      50      Alignments to save      0
Optimized scores to save      0      Display context      0

```

SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              1          3          1.07

```

```

Times:      CPU      Total Elapsed
            00:03:25.11      00:07:03.00

```

```

Number of residues:      25433612
Number of sequences searched:      94553
Number of scores above cutoff:      3985

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 4 standard deviations above mean ****						
1. YCA8_YEAST	HYPOTHETICAL 13.1 KD PROTEIN	119	6	6	4.69	0
2. S19410	Hypothetical protein YOL80 -	119	6	6	4.69	0
3. S17923	rbcl intron protein - Green a	274	6	6	4.69	0
4. MTD1_DESDN	MODIFICATION METHYLASE DDEI (416	6	7	4.69	0
5. S00543	Site-specific methyltransfera	416	6	7	4.69	0
6. AMYP_RAT	ALPHA-AMYLASE PANCREATIC PREC	503	6	6	4.69	0
7. ALRTP	alpha-Amylase precursor, panc	503	6	6	4.69	0
8. P50263	Sequence encoded by pancreati	507	6	6	4.69	0
9. ALMSPC	alpha-Amylase D(C) precursor,	508	6	6	4.69	0
10. ALMSP1	alpha-Amylase A1 precursor, p	508	6	6	4.69	0
11. AMYS_MOUSE	ALPHA-AMYLASE SALIVARY AND HE	511	6	6	4.69	0
12. ALMS5	alpha-Amylase precursor, full	511	6	6	4.69	0
13. AVR1_MOUSE	ACTIVIN RECEPTOR TYPE I PROTE	713	6	6	4.69	0
14. AVR1_HUMAN	ACTIVIN RECEPTOR TYPE I PROTE	713	6	6	4.69	0
15. S7942						

17.	S22345	*Activin receptor - Human	513	6	6	4.69	0
18.	J01486	Activin receptor II precursor	513	6	6	4.69	0
19.	A37913	*mkl1 protein - Yeast (Schizo	581	6	6	4.69	0
20.	SWI5_YEAST	TRANSCRIPTIONAL FACTOR SWI5.	709	6	6	4.69	0
21.	TWBY55	Transcription factor SWI5 - Y	709	6	6	4.69	0
22.	S18512	Cell division control protein	1035	6	6	4.69	0
23.	POLG_POL1M	GENOME POLYPROTEIN (COAT PROT	2207	6	6	4.69	0
24.	S09553	Genome polyprotein - Human po	2207	6	6	4.69	0
25.	GNNY1P	Genome polyprotein (version 1	2207	6	6	4.69	0
26.	POLH_POL1M	GENOME POLYPROTEIN (COAT PROT	2209	6	6	4.69	0
27.	POLG_POL1S	GENOME POLYPROTEIN (COAT PROT	2209	6	6	4.69	0
28.	GNNY3P	Genome polyprotein - Human po	2209	6	6	4.69	0
29.	GNNY2P	Genome polyprotein (version 2	2209	6	6	4.69	0
30.	A37113	*Ryanodine receptor, cardiac	4969	6	6	4.69	0
31.	R10834	Rianodin receptor.	4987	6	6	4.69	0
32.	RYNR_HUMAN	RYANODINE RECEPTOR, SKELETAL	5032	6	6	4.69	0
33.	A35041	*Ryanodine receptor - Human	5032	6	6	4.69	0
34.	B35041	*Ryanodine receptor - Rabbit	5034	6	6	4.69	0
35.	S18135	*Calcium release channel - Pi	5034	6	6	4.69	0
36.	RYNR_RABIT	RYANODINE RECEPTOR, SKELETAL	5037	6	6	4.69	0
37.	S04654	Ryanodine receptor - Rabbit	5037	6	6	4.69	0
38.	R11510	Ryanodine receptor deduced fr	5072	6	6	4.69	0
**** 3 standard deviations above mean ****							
39.	R04972	Papilloma virus type 16 L1 pe	20	5	5	3.75	0
40.	P81545	Human insulin acceptor protei	33	5	5	3.75	0
41.	A36357	*Cytochrome-c oxidase chain I	66	5	5	3.75	0
42.	NXL3_DENPO	LONG NEUROTOXIN 3 (TOXIN VN2)	72	5	5	3.75	0
43.	NXL2_DENPO	LONG NEUROTOXIN 2 (NEUROTOXIN	72	5	5	3.75	0
44.	NXL1_DENPO	LONG NEUROTOXIN 1 (NEUROTOXIN	72	5	5	3.75	0
45.	N2EP2D	Long neurotoxin 2 - Black mam	72	5	5	3.75	0
46.	N2EP1D	Long neurotoxin 1 - Black mam	72	5	5	3.75	0
47.	A35991	*Retinoic acid receptor gamma	74	5	5	3.75	0
48.	THIO_BPT4	THIOREDOXIN.	87	5	5	3.75	0
49.	G30292	*Gene nrdC protein - Phage T4	87	5	5	3.75	0
50.	TXBPT4	Thioredoxin - Phage T4	87	5	5	3.75	0

No alignments saved.

> U <
0| |0 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-8.res made by alexk on Thu 25 Feb 93 10:24:25-PST.

Query sequence being compared: CELSA-8 (1-11)
Number of sequences searched: 94553
Number of scores above cutoff: 3979

Results of the initial comparison of CELSA-8 (1-11) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries

100000-

-

N - *

U50000-

M -

B -

E - *

R -

* -

D -

F10000-

-

S -

E 5000- *

Q -

U -

E -

N -

C -

E -

S 1000-

-

-

500-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

-

Best Available Copy

SCORE	0	1	2	3	4	5	6	7
STDEV	0	1	2	3	4	5	6	7

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	3	1.04

Times:	CPU	Total Elapsed
	00:03:41.11	00:08:23.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	3979

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. FAS_CHICK	FATTY ACID SYNTHASE (EC 2.3.1	2446	7	7	5.77	0
2. XYCHFA	Fatty-acid synthase - Chicken	2446	7	7	5.77	0
**** 4 standard deviations above mean ****						
3. P70030	Secretory signal sequence of	26	6	6	4.81	0
4. P70033	Secretory signal sequence of	39	6	6	4.81	0
5. P70029	Secretory signal sequence of	50	6	6	4.81	0
6. S17024	Hypothetical protein (IFM1 3'	124	6	6	4.81	0
7. YIFM_YEAST	HYPOTHETICAL PROTEIN IN IFM1	125	6	6	4.81	0
8. S20178	*Hypothetical protein - Yeast	125	6	6	4.81	0
9. FMC1_ECOLI	CFA/I FIMBRIAL SUBUNIT B PREC	170	6	6	4.81	0
10. YQECC1	CFA1 fimbrial protein precurs	170	6	6	4.81	0
11. SSBR_ECOLI	SINGLE-STRAND BINDING PROTEIN	174	6	6	4.81	0
12. SSBP_ECOLI	SINGLE-STRAND BINDING PROTEIN	174	6	6	4.81	0
13. A38487	*Helix-destabilizing protein	174	6	6	4.81	0
14. DNECIR	Helix-destabilizing protein -	175	6	6	4.81	0

16. IGF3_HUMAN	INSULIN-LIKE GROWTH FACTOR BI	291	6	6	4.81	0
17. A34651	*Insulin-like growth factor b	291	6	6	4.81	0
18. IOHU3	Insulin-like growth factor bi	291	6	6	4.81	0
19. R05596	Somatomedin carrier protein s	291	6	6	4.81	0
20. P92300	Sequence of human insulin-lik	291	6	6	4.81	0
21. IBP3_RAT	INSULIN-LIKE GROWTH FACTOR BI	292	6	6	4.81	0
22. A36748	*Insulin-like growth factor b	292	6	6	4.81	0
23. S11738	Hemagglutinin precursor - Inf	378	6	7	4.81	0
24. A39314	*Gastricsin precursor - Bullf	384	6	6	4.81	0
25. PEPC_HUMAN	PROGASTRICSIN PRECURSOR (EC 3	388	6	6	4.81	0
26. A29937	*Gastricsin precursor - Human	388	6	6	4.81	0
27. A31811	*Gastricsin precursor - Human	388	6	6	4.81	0
28. PEPC_RAT	GASTRICSIN PRECURSOR (EC 3.4.	392	6	6	4.81	0
29. A33510	*Gastricsin - Rat #EC-number	392	6	6	4.81	0
30. A24608	Pepsin A precursor - Rat #EC-	392	6	6	4.81	0
31. YAJA_ECOLI	HYPOTHETICAL 44.7 KD PROTEIN	400	6	7	4.81	0
32. JS0349	Hypothetical 45K protein (sbc	400	6	7	4.81	0
33. CYAA_TRYEQ	ADENYLATE CYCLASE (EC 4.6.1.1	469	6	6	4.81	0
34. S16359	Adenylate cyclase - Trypanoso	469	6	6	4.81	0
35. DM6E_CHLTR	60 KD OUTER MEMBRANE PROTEIN	547	6	6	4.81	0
36. S13120	*Omp2 protein - Chlamydia tra	547	6	6	4.81	0
37. PC1_MOUSE	PLASMA-CELL MEMBRANE GLYCOPRO	871	6	6	4.81	0
38. A27410	Plasma cell membrane protein	905	6	6	4.81	0
39. CARB_BACSU	CARBAMOYL-PHOSPHATE SYNTHASE	1071	6	6	4.81	0
40. F39845	*Carbamyl phosphate synthase	1071	6	6	4.81	0
**** 3 standard deviations above mean ****						
41. P98464	Sequence of C. trachomatis se	14	5	5	3.85	0
42. P98452	Sequence of C. trachomatis se	14	5	5	3.85	0
43. P98448	Sequence of C. trachomatis se	14	5	5	3.85	0
44. P98444	Sequence of C. trachomatis se	14	5	5	3.85	0
45. P98468	Sequence of C. trachomatis se	14	5	5	3.85	0
46. P98460	Sequence of C. trachomatis se	14	5	5	3.85	0
47. YPS1_PLEBO	HYPOTHETICAL 4.6 KD PROTEIN (40	5	5	3.85	0
48. PSEJ_CYAPA	PHOTOSYSTEM II REACTION CENTR	40	5	6	3.85	0
49. S08038	Hypothetical protein 1 - Plec	40	5	5	3.85	0
50. PQ0104	Microbial serine proteinase I	41	5	5	3.85	0

No alignments saved.

< 0 >
01 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-9.res made by alexk on Thu 25 Feb 93 10:39:58-PST.

Query sequence being compared: CELSA-9 (1-16)
Number of sequences searched: 94553
Number of scores above cutoff: 4422

Results of the initial comparison of CELSA-9 (1-16) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries

100000-
-
N
-
U50000-
M
-
B
-
E
-
R
-
D
-
F10000-
-
S
-
E 5000-
Q
-
U
-
E
-
N
-
C
-
E
-
S 1000-
-
-
500-
-
-
-
-
100-
-
-
50-
-
-
-
-
10-
-
5-
-
-

*

*

*

*

*

*

*

SCORE	0	1	2	3	4	6	7	8	9	10
STDEV	-1	0	1	2	3	4	5	6	7	

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.06

Times:	CPU	Total Elapsed
	00:03:26.02	00:06:54.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4422

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 7 standard deviations above mean ****						
1. PC1_HUMAN	PLASMA-CELL MEMBRANE GLYCOPRO	873	10	11	7.58	0
2. S21706	*Pyrophosphatase - Human	925	10	11	7.58	0
3. A39216	*Plasma cell membrane protein	925	10	11	7.58	0
**** 4 standard deviations above mean ****						
4. VD04_FOWP1	25.6 KD PROTEIN.	218	7	8	4.74	0
5. A35216	*FPD4 protein - Fowlpox virus	218	7	8	4.74	0
6. R21437	PE-40 protein contg. a methio	361	7	7	4.74	0
7. R21436	PE-40 somatostatin substitute	361	7	7	4.74	0
8. R21435	PE-40 somatostatin substitute	361	7	7	4.74	0
9. R20201	TGF-alpha-PE40ab.	419	7	7	4.74	0
10. R20199	TGF-alpha-PE40aB.	419	7	7	4.74	0
11. R07054	PE40AB protein comprising a p	419	7	7	4.74	0
12. R20200	TGF-alpha-PE40Ab.	420	7	7	4.74	0
13. R06450	TGF-alpha-PE40-ab modified ps	420	7	7	4.74	0
14. R06449	TGF-alpha-PE40-ab modified ps	420	7	7	4.74	0

16. R06447	TGF-57-Pseudomonas exotoxin 4	420	7	7	4.74	0
17. R06994	PE40ab protein comprising a p	420	7	7	4.74	0
18. R06992	PE40aB protein comprising a p	420	7	7	4.74	0
19. R06993	PE40Ab protein comprising a p	420	7	7	4.74	0
20. AMPL_BOVIN	LEUCINE AMINOPEPTIDASE (EC 3.	478	7	7	4.74	0
21. APBOL	Cytosol aminopeptidase - Bovi	478	7	7	4.74	0
22. R04934	Immunotoxin hybrid of human i	496	7	7	4.74	0
23. R04920	Immunoprotein PEX46.	549	7	7	4.74	0
24. R04923	Immunoprotein TANG11.	557	7	7	4.74	0
25. R04919	Immunoprotein PEX45.	574	7	7	4.74	0
26. R04924	Immunoprotein TANG12.	577	7	7	4.74	0
27. TDXA_PSEAE	EXOTOXIN A PRECURSOR (NAD-DEP	638	7	7	4.74	0
28. A30347	*Exotoxin A precursor - Pseud	638	7	7	4.74	0
29. FEPA_ECOLI	FERRIC ENTEROCHELIN RECEPTOR	745	7	7	4.74	0
30. QRECFC	Ferrienterochelin receptor pr	745	7	7	4.74	0
**** 3 standard deviations above mean ****						
31. PN0088	*Matrix protein M3 - Avian in	68	6	7	3.79	0
32. PN0085	*Matrix protein M3 - Avian in	68	6	7	3.79	0
33. ISS_ECOLI	HYPOTHETICAL ISS PROTEIN.	102	6	6	3.79	0
34. JE0036	Hypothetical iss protein - Es	102	6	6	3.79	0
35. ALB2_PEA	ALBUMIN 2 (PA2).	231	6	6	3.79	0
36. S06248	Albumin 2 - Garden pea	231	6	6	3.79	0
37. JQ0319	Hypothetical 27K protein - Xa	261	6	6	3.79	0
38. ETB_STAAU	EXFOLIATIVE TOXIN B PRECURSOR	277	6	7	3.79	0
39. PRSAEB	Epidermolytic toxin B precurs	277	6	7	3.79	0
40. MIM1_CHICK	MYELOID PROTEIN-1 PRECURSOR.	326	6	6	3.79	0
41. A33755	*myb-induced myeloid protein	326	6	6	3.79	0
42. BYR1_SCHPO	PROTEIN KINASE BYR1 (EC 2.7.1	340	6	6	3.79	0
43. OKBYR1	Protein kinase byr1 - Yeast (340	6	6	3.79	0
44. YPIX_CLOPE	HYPOTHETICAL 38.4 KD PROTEIN	342	6	7	3.79	0
45. JT0370	Hypothetical protein g10 - Cl	342	6	7	3.79	0
46. VMAT_SYNV	MATRIX PROTEIN (M2 PROTEIN).	345	6	7	3.79	0
47. MFVNSY	Matrix protein - Sonchus yell	345	6	7	3.79	0
48. K1CR_MOUSE	KERATIN, TYPE I CYTOSKELETAL	422	6	7	3.79	0
49. JT0406	Keratin, type 1 cytoskeletal	423	6	7	3.79	0
50. A25621	Endo B cytokeratin - Mouse	423	6	7	3.79	0

No alignments saved.

Best Available Copy

0-----

SCORE	0	1	2	3	4	5	6	7
STDEV	-1	0	1	2	3	4	5	6

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.10

Times:	CPU	Total Elapsed
	00:03:25.96	00:07:28.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4430

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 4 standard deviations above mean ****						
1. WMBEA1	Ribonucleoside-diphosphate re	321	7	7	4.56	0
2. A36062	*Catalase - Maize #EC-number	491	7	8	4.56	0
3. S18819	*Catalase - Maize #EC-number	491	7	8	4.56	0
4. CATA_IPOBA	CATALASE (EC 1.11.1.6).	492	7	8	4.56	0
5. CAT1_GOSHI	CATALASE ISOZYME A (EC 1.11.1	492	7	8	4.56	0
6. S07124	*Catalase - Sweet potato #EC-	492	7	8	4.56	0
7. S20999	*Catalase - Soybean #EC-numbe	492	7	8	4.56	0
8. S10770	*Catalase - Upland cotton #EC	492	7	8	4.56	0
9. S10395	Catalase chain 1 - Upland cot	492	7	8	4.56	0
10. CATA_PEA	CATALASE (EC 1.11.1.6).	494	7	8	4.56	0
11. S18346	Catalase - Garden pea #EC-num	494	7	8	4.56	0
**** 3 standard deviations above mean ****						
12. C20554	Hemocyanin LpIIa - Atlantic h	20	6	6	3.65	0
13. A20554	Hemocyanin LpI chain - Atlant	24	6	6	3.65	0
14. F20554	Hemocyanin LpIV - Atlantic h	24	6	6	3.65	0

16. SC1037	*DNA gyrase chain B - Lyme di	84	6	6	3.65	0
17. A28002	Apolipoprotein B-48 - Human (106	6	6	3.65	0
18. JN0145	Hypothetical 13.6K protein (d	117	6	7	3.65	0
19. HEM4_BACSU	PUTATIVE UROPORPHYRINOGEN-III	130	6	6	3.65	0
20. D35252	*Putative hemD protein - Baci	130	6	6	3.65	0
21. P82320	PAP-III isolated from biologi	145	6	6	3.65	0
22. JT0961	Glutathione synthase large ch	285	6	6	3.65	0
23. CENA_MOUSE	CENTROSOMIN A.	289	6	6	3.65	0
24. S13800	*Centrosomin A - Mouse	289	6	6	3.65	0
25. R22403	Partial sequence of N-lipocor	304	6	6	3.65	0
26. B41002	Annexin II (clones E4 and F4)	314	6	6	3.65	0
27. R10689	Cephalosporin antibiotic bios	319	6	6	3.65	0
28. ANX3_HUMAN	ANNEXIN III (LIPOCORTIN III)	323	6	6	3.65	0
29. LUHU3	Annexin III - Human	323	6	6	3.65	0
30. P91362	Human lipocortin-III.	323	6	6	3.65	0
31. ANX3_RAT	ANNEXIN III (LIPOCORTIN III)	324	6	6	3.65	0
32. LURT3	Annexin III - Rat	324	6	6	3.65	0
33. VG16_BPPZA	ENCAPSIDATION PROTEIN (LATE P	332	6	6	3.65	0
34. VG16_BPPH2	ENCAPSIDATION PROTEIN (LATE P	332	6	6	3.65	0
35. JQ0168	Gene 16 protein - Phage phi-2	332	6	6	3.65	0
36. WMBP26	Gene 16 protein - Phage phi-2	332	6	6	3.65	0
37. WMBP16	Gene 16 protein - Phage PZA	332	6	6	3.65	0
38. RIR2_HSV23	RIBONUCLEOSIDE-DIPHOSPHATE RE	337	6	6	3.65	0
39. WMBE2	Ribonucleoside-diphosphate re	337	6	6	3.65	0
40. WMBE32	Ribonucleoside-diphosphate re	337	6	6	3.65	0
41. ANX2_MOUSE	ANNEXIN II (LIPOCORTIN II) (C	338	6	6	3.65	0
42. ANX2_HUMAN	ANNEXIN II (LIPOCORTIN II) (C	338	6	6	3.65	0
43. ANX2_CHICK	ANNEXIN II (LIPOCORTIN II) (C	338	6	6	3.65	0
44. ANX2_BOVIN	ANNEXIN II (LIPOCORTIN II) (C	338	6	6	3.65	0
45. RIR2_HSV1K	RIBONUCLEOSIDE-DIPHOSPHATE RE	339	6	6	3.65	0
46. ANXB_XENLA	ANNEXIN II TYPE I (LIPOCORTIN	339	6	6	3.65	0
47. ANX2_XENLA	ANNEXIN II TYPE II (LIPOCORTI	339	6	6	3.65	0
48. LUCH2	Annexin II - Chicken	339	6	6	3.65	0
49. LUMS36	Annexin II - Mouse	339	6	6	3.65	0
50. LUBQ36	Annexin II - Bovine	339	6	6	3.65	0

No alignments saved.

0110 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-11.res made by alexk on Thu 25 Feb 93 10:24:30-PST.

Query sequence being compared: CELSA-11 (1-23)
Number of sequences searched: 94553
Number of scores above cutoff: 3948

Results of the initial comparison of CELSA-11 (1-23) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries

100000-
-
N
-
U50000-
M
-
B
-
E
-
R
-
-
O
-
F10000-
*
S
-
E 5000-
Q
-
U
-
E
-
N
-
C
-
E
-
S 1000-
-
-
500-
-
-
-
-
100-
-
-
50-
-
-
-
-
10-
-
5-
-
-

* *
*
*
*
*
*
*

Best Available Copy

SCORE	0	1	2	3	4	5	6	7	8	9
STDEV	-1	0	1	2	3	4	5	6	7	8

PARAMETERS

Similarity matrix Unitary K-tuple 2
 Mismatch penalty 5 Joining penalty 20
 Gap penalty 1.00 Window size 5
 Gap size penalty 0.26
 Cutoff score 0
 Randomization group 0

Initial scores to save 50 Alignments to save 0
 Optimized scores to save 0 Display context 0

SEARCH STATISTICS

Scores: Mean Median Standard Deviation
 2 4 1.13

Times: CPU Total Elapsed
 00:03:59.10 00:08:28.00

Number of residues: 25433612
 Number of sequences searched: 94553
 Number of scores above cutoff: 3948

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.
 Cut-off raised to 6.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 6 standard deviations above mean ****						
1. EFTU_EUGGR	ELONGATION FACTOR TU (EF-TU).	409	9	10	6.18	0
2. S02254	Elongation factor Tu - Euglen	409	9	10	6.18	0
3. EFEGT	Elongation factor Tu - Euglen	409	9	10	6.18	0
**** 5 standard deviations above mean ****						
4. XYLR_STAXY	XYLOSE REPRESSOR.	383	8	8	5.29	0
5. S16529	*xylR protein - Staphylococcu	383	8	8	5.29	0
**** 4 standard deviations above mean ****						
6. YDP3_LACLA	HYPOTHETICAL 18.0 KD PROTEIN	161	7	10	4.41	0
7. B33374	*Hypothetical protein BVRF1 -	193	7	9	4.41	0
8. YEIB_ECOLI	HYPOTHETICAL PROTEIN IN GAL5	196	7	7	4.41	0
9. S19934	Hypothetical protein - Escher	196	7	7	4.41	0
10. UL52_HSVSA	HYPOTHETICAL BSLF1 PROTEIN HO	205	7	7	4.41	0
11. Q0BEHA	BSLF1 protein - Saimirine her	205	7	7	4.41	0
12. KAD_PARDF	ADENYLATE KINASE (EC 2.7.4.7)	217	7	7	4.41	0

14. NIPC	Adenylate kinase - Paracoccus	217	7	7	4.41	0
15. KAD1_YEAST	ADENYLATE KINASE CYTOSOLIC (E	222	7	8	4.41	0
16. KIBYA	Adenylate kinase - Yeast (Sac	222	7	8	4.41	0
17. R03339	VP1 sequence for HRV serotype	287	7	9	4.41	0
18. R03340	VP1 sequence for HRV serotype	291	7	8	4.41	0
19. CMG1_BACSU	CMG OPERON PROTEIN 1.	356	7	7	4.41	0
20. B30338	*comG operon protein 1 - Baci	356	7	7	4.41	0
21. RRPD_LYCVW	RNA POLYMERASE (EC 2.7.7.48)	363	7	9	4.41	0
22. DDLA_ECOLI	D-ALANINE--D-ALANINE LIGASE A	364	7	7	4.41	0
23. CEECD	D-Alanine--D-alanine ligase A	364	7	7	4.41	0
24. TRPB_ACICA	TRYPTOPHAN SYNTHASE BETA CHAI	403	7	8	4.41	0
25. B36151	*Tryptophan synthase beta cha	403	7	8	4.41	0
26. FIBG_BOVIN	FIBRINOGEN GAMMA-B CHAIN PREC	444	7	9	4.41	0
27. S05313	Fibrinogen gamma-B chain prec	444	7	9	4.41	0
28. FUCD_DICDI	ALPHA-L-FUCOSIDASE PRECURSOR	461	7	10	4.41	0
29. A30364	alpha-L-Fucosidase homolog pr	461	7	10	4.41	0
30. A41533	*Reticuline oxidase precursor	538	7	7	4.41	0
31. VGNZPD	Fusion glycoprotein precursor	631	7	8	4.41	0
32. NUSM_ASPNI	NADH-UBIQUINONE OXIDOREDUCTAS	657	7	8	4.41	0
33. S04724	*NADH dehydrogenase (ubiquino	657	7	8	4.41	0
34. UL15_HSV6U	HYPOTHETICAL PROTEIN 12L (ORF	667	7	9	4.41	0
35. Q0BEH6	12L protein - Human herpesvir	667	7	9	4.41	0
36. CAP1_HUMAN	CALPAIN I, LARGE (CATALYTIC)	714	7	9	4.41	0
37. S10591	*Cysteine protease - Human	714	7	9	4.41	0
38. CIHUH	Calpain I heavy chain - Human	714	7	9	4.41	0
39. POLG_HRV1A	GENOME POLYPROTEIN (COAT PROT	832	7	9	4.41	0
40. VCAP_HSV11	MAJOR CAPSID PROTEIN (MCP).	1374	7	7	4.41	0
41. A30084	*Gene UL19 protein (major cap	1374	7	7	4.41	0
42. VCB17	Major capsid protein - Human	1374	7	7	4.41	0
43. ARO1_YEAST	PENTAFUNCTIONAL AROM POLYPEPT	1588	7	8	4.41	0
44. BVBYA1	ARO1 protein - Yeast (Sacchar	1588	7	8	4.41	0
45. S18644	*Multifunctional aminoacyl-tR	1714	7	8	4.41	0
46. RRPD_TACV	RNA POLYMERASE (EC 2.7.7.48).	2210	7	8	4.41	0
47. RRXPTV	RNA-directed RNA polymerase -	2210	7	8	4.41	0
48. S22011	*6-Deoxyerythronolide B synth	3567	7	8	4.41	0
49. A41819	*637K proline-rich peptide pr	5762	7	9	4.41	0
**** 3 standard deviations above mean ****						
50. R15527	Immunopeptide derived from HP	20	6	6	3.53	0

No alignments saved.

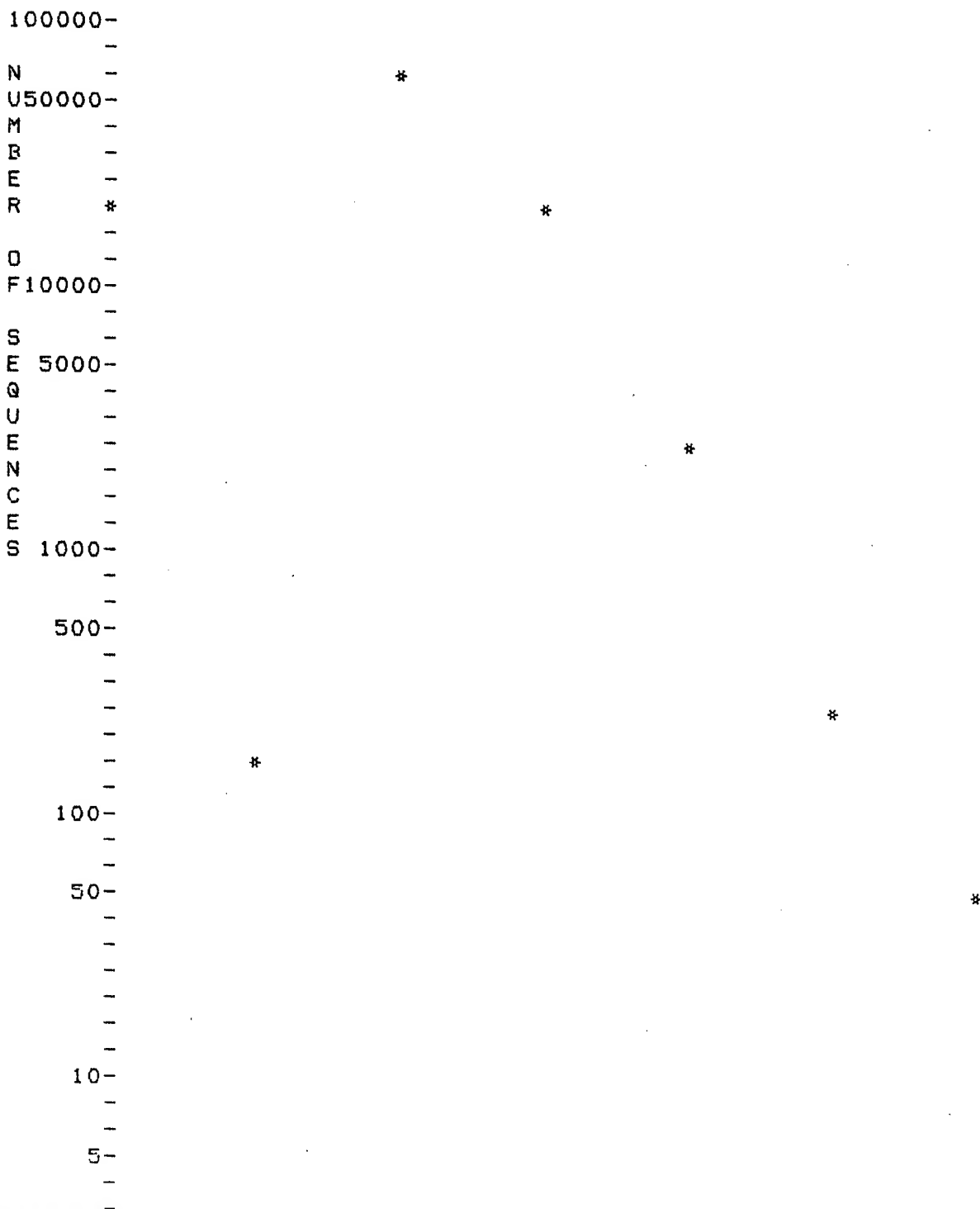
< 0 >
0| 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-12.res made by alexk on Thu 25 Feb 93 10:32:32-PST.

Query sequence being compared: CELSA-12 (1-9)
Number of sequences searched: 94553
Number of scores above cutoff: 3845

Results of the initial comparison of CELSA-12 (1-9) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

```

0-----
SCORE 0| 1| 2| 3| 4| 5| 6| 7
STDEV 0| 1| 2| 3| 4| 5| 6| 7

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5      Joining penalty      20
Gap penalty      1.00      Window size      5
Gap size penalty      0.26
Cutoff score      0
Randomization group      0

Initial scores to save      50      Alignments to save      0
Optimized scores to save      0      Display context      0

```

SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              1          3          1.04

```

```

Times:      CPU      Total Elapsed
            00:03:23.11      00:07:28.00

```

```

Number of residues:      25433612
Number of sequences searched:      94553
Number of scores above cutoff:      3845

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. CYB_DENBE	CYTOCHROME B (EC 1.10.2.2).	394	7	7	5.76	0
2. S20141	Cytochrome b - Evening primro	394	7	7	5.76	0
3. CBOBE	Cytochrome b - Evening primro	394	7	7	5.76	0
**** 4 standard deviations above mean ****						
4. R13364	P691 HCV antigen (691-714).	24	6	6	4.80	0
5. P92042	Sequence encoded in the hepat	141	6	6	4.80	0
6. P90159	Sequence of hepatitis C virus	141	6	6	4.80	0
7. S21305	*Apocytochrome b - Rice	334	6	6	4.80	0
8. CYB_CHLSM	CYTOCHROME B (EC 1.10.2.2).	381	6	6	4.80	0
9. CYB_CHLRE	CYTOCHROME B (EC 1.10.2.2).	381	6	6	4.80	0
10. S12023	*Cytochrome b - Chlamydomonas	381	6	6	4.80	0
11. S11966	*Cytochrome b - Chlamydomonas	381	6	6	4.80	0
12. S12016	*Cytochrome b - Chlamydomonas	381	6	6	4.80	0
13. CYB_MAIZE	CYTOCHROME B (EC 1.10.2.2).	388	6	6	4.80	0
14. CBZM	Cytochrome b - Maize mitochondr	388	6	6	4.80	0

16. CYB	Cytochrome b - Fava bean mito	392	6	6	4.80	0
17. S17427	*Cytochrome b - Potato mitoch	393	6	6	4.80	0
18. CYB_DRYSA	CYTOCHROME B (EC 1.10.2.2).	397	6	6	4.80	0
19. S20659	*Apocytochrome b - Rice	397	6	6	4.80	0
20. CBRZ	Cytochrome b - Rice mitochond	397	6	6	4.80	0
21. CYB_WHEAT	CYTOCHROME B (EC 1.10.2.2).	398	6	6	4.80	0
22. A22931	Cytochrome b - Wheat mitochon	398	6	6	4.80	0
23. CYB_MARPO	CYTOCHROME B (EC 1.10.2.2).	404	6	6	4.80	0
24. CARA_YEAST	CARBAMOYL-PHOSPHATE SYNTHASE,	411	6	6	4.80	0
25. SYBYCS	Carbamoyl-phosphate synthase	411	6	6	4.80	0
26. S20660	*Pseudo apocytochrome b - Ric	449	6	6	4.80	0
27. P92049	Sequence encoded by segment o	454	6	6	4.80	0
28. P90183	Sequence of hepatitis C virus	454	6	6	4.80	0
29. R21577	HCV CKS-NS1 - pHCV-107.	467	6	6	4.80	0
30. JQ1366	Polyprotein - Hepatitis C vir	716	6	6	4.80	0
31. YCD9_YEAST	HYPOTHETICAL 86.0 KD PROTEIN	759	6	6	4.80	0
32. S19367	Hypothetical protein YCL39W -	759	6	6	4.80	0
33. P90164	Peptide encoded by composite	2261	6	6	4.80	0
34. P92047	Sequence encoded in the hepat	2301	6	6	4.80	0
35. P92050	Sequence encoded in the hepat	2436	6	6	4.80	0
36. P90288	Peptide encoded by composite	2462	6	6	4.80	0
37. R08123	Hepatitis C virus polypeptide	2772	6	6	4.80	0
38. R24440	Composite HCV HC-J1/CDC/CHI p	2894	6	6	4.80	0
39. R08124	Hepatitis C virus putative po	2955	6	6	4.80	0
40. POLG_HCVH	GENOME POLYPROTEIN (CONTAINS:	3011	6	6	4.80	0
41. POLG_HCV1	GENOME POLYPROTEIN (CONTAINS:	3011	6	6	4.80	0
42. GNWVC3	Genome polyprotein - Hepatiti	3011	6	6	4.80	0
43. R22154	NANBV Hutch c59 isolate genom	3011	6	6	4.80	0
44. R21519	Compiled HCV sequence.	3011	6	6	4.80	0
**** 3 standard deviations above mean ****						
45. R15366	Ig idiotypic determinant PSL3	12	5	5	3.84	0
46. R11408	Hepatitis B virus pre-S2 QE2-	38	5	5	3.84	0
47. A24265	Phosphate transport protein,	47	5	5	3.84	0
48. A25521	Ig kappa chain V region (321)	54	5	6	3.84	0
49. S19988	*Hypothetical protein 1708 -	65	5	5	3.84	0
50. S19981	*Hypothetical protein 1708 -	65	5	5	3.84	0

No alignments saved.

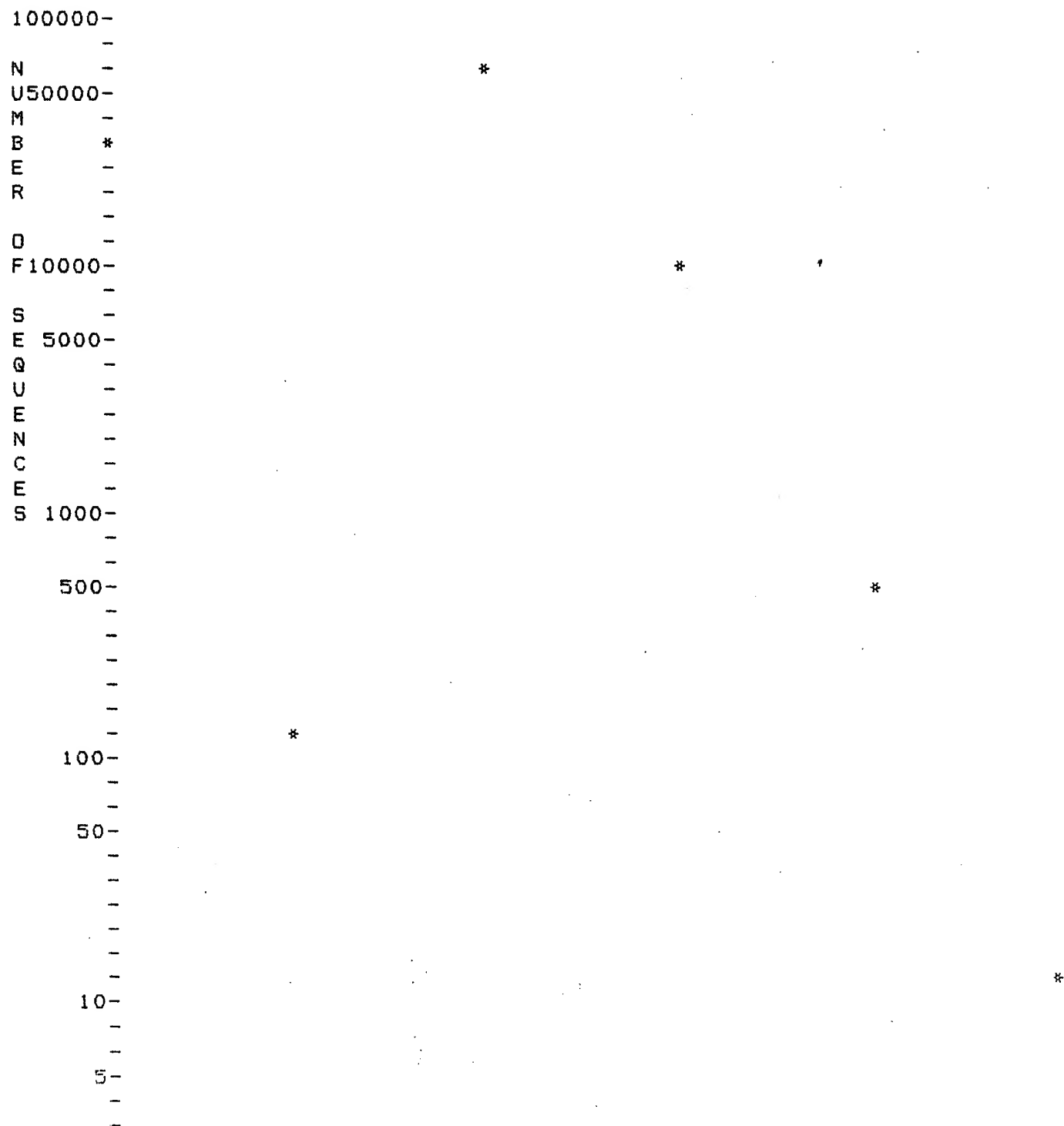
> 0 <
0| 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-13.res made by alexk on Thu 25 Feb 93 10:39:48-PST.

Query sequence being compared: CELSA-13 (1-5)
Number of sequences searched: 94553
Number of scores above cutoff: 4106

Results of the initial comparison of CELSA-13 (1-5) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

0-----

SCORE	0	1	1	2	2	3	3	4	4	5
STDEV	0		1		2		3		4	5

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	1	3	1.01

Times:	CPU	Total Elapsed
	00:03:18.07	00:06:44.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4106

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.

The scores below are sorted by initial score.
Significance is calculated based on initial score.

11 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. A37113	*Ryanodine receptor, cardiac	4969	5	5	3.98	0
2. BVFFSL	sol protein, large splice for	1597	5	5	3.98	0
3. SOL_DROME	SMALL OPTIC LOBES PROTEIN.	1597	5	5	3.98	0
4. CP51_CANTR	CYTOCHROME P450 L1 (P450-L1A1	528	5	5	3.98	0
5. A31854	Cytochrome P450 51 lanosterol	528	5	5	3.98	0
6. S12042	*Sugar transport protein STP1	522	5	5	3.98	0
7. STP1_ARATH	GLUCOSE TRANSPORTER (SUGAR CA	522	5	5	3.98	0
8. S14627	*Glucose transport protein -	522	5	5	3.98	0
9. E35901	*Calcium channel alpha-1 chai	164	5	5	3.98	0
10. J00700	Hypothetical 11K protein (mmo	103	5	5	3.98	0
11. YMMO_METCA	HYPOTHETICAL 11.9 KD PROTEIN	103	5	5	3.98	0

The list of other best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 2 standard deviations above mean ****						
12. C04775	putative transmembrane	22	0	4	0.00	0

14. D24735	Glutathione transferase, 2-2	17	4	4	2.98	0
15. A24735	Glutathione transferase, 1-1	26	4	4	2.98	0
16. S21278	*Glutathione transferase chai	28	4	4	2.98	0
17. S03358	*Glutathione transferase - Ra	31	4	4	2.98	0
18. S09585	*Glutathione transferase - Ra	31	4	4	2.98	0
19. V07K_PMV	7 KD PROTEIN (ORF 4).	68	4	4	2.98	0
20. J00099	Hypothetical 7K protein - Pap	68	4	4	2.98	0
21. P81053	Sequence of rhinovirus HRV2 v	69	4	4	2.98	0
22. P81052	Sequence of rhinovirus HRV89	69	4	4	2.98	0
23. TRAL_ECOLI	TRAL PROTEIN.	91	4	4	2.98	0
24. BVECTL	traL protein - Escherichia co	91	4	4	2.98	0
25. S03311	traL protein - Escherichia co	92	4	4	2.98	0
26. VES_HP42	PROBABLE E5 PROTEIN.	95	4	4	2.98	0
27. W5WL42	E5 protein - Human papillomav	95	4	4	2.98	0
28. YR12_CYAPA	HYPOTHETICAL 12.3 KD PROTEIN	102	4	4	2.98	0
29. S12809	Hypothetical protein (rpl3 5'	102	4	4	2.98	0
30. S14355	*Glutathione transferase - Ra	113	4	4	2.98	0
31. THIO_ANASP	THIOREDOXIN.	115	4	4	2.98	0
32. A32233	*Thioredoxin - Anabaena sp.	115	4	4	2.98	0
33. VG67_BPPZA	EARLY PROTEIN GP16.7.	130	4	4	2.98	0
34. VG67_BPPH5	EARLY PROTEIN GP16.7.	130	4	4	2.98	0
35. VG67_BPPH2	EARLY PROTEIN GP16.7.	130	4	4	2.98	0
36. JN0033	Early protein gp15 - Phage ph	130	4	4	2.98	0
37. WRBPF5	Early protein gp15 - Phage ph	130	4	4	2.98	0
38. WRBP67	Early protein gp15 - Phage P2	130	4	4	2.98	0
39. YKP7_KLULA	HYPOTHETICAL KILLER PLASMID P	132	4	4	2.98	0
40. S15966	*Hypothetical protein 7 - Yea	132	4	4	2.98	0
41. S00965	Hypothetical protein 7 - Yeas	132	4	4	2.98	0
42. S19269	*Glutathione transferase Yc2	139	4	4	2.98	0
43. BCP_ECOLI	BACTERIOFERRITIN COMIGRATORY	156	4	4	2.98	0
44. V02_FMVD	PROTEIN 2.	164	4	4	2.98	0
45. S01280	Hypothetical protein 2 - Figw	164	4	4	2.98	0
46. NU6M_MOUSE	NADH-UBIQUINONE OXIDOREDUCTAS	172	4	4	2.98	0
47. DEMSN6	NADH dehydrogenase (ubiquinon	172	4	4	2.98	0
48. PSP_BOVIN	PANCREATIC STONE PROTEIN PREC	175	4	4	2.98	0
49. A37194	*Pancreatic thread protein pr	175	4	4	2.98	0
50. JS0679	Hypothetical 20K protein - Hu	175	4	4	2.98	0

No alignments saved.

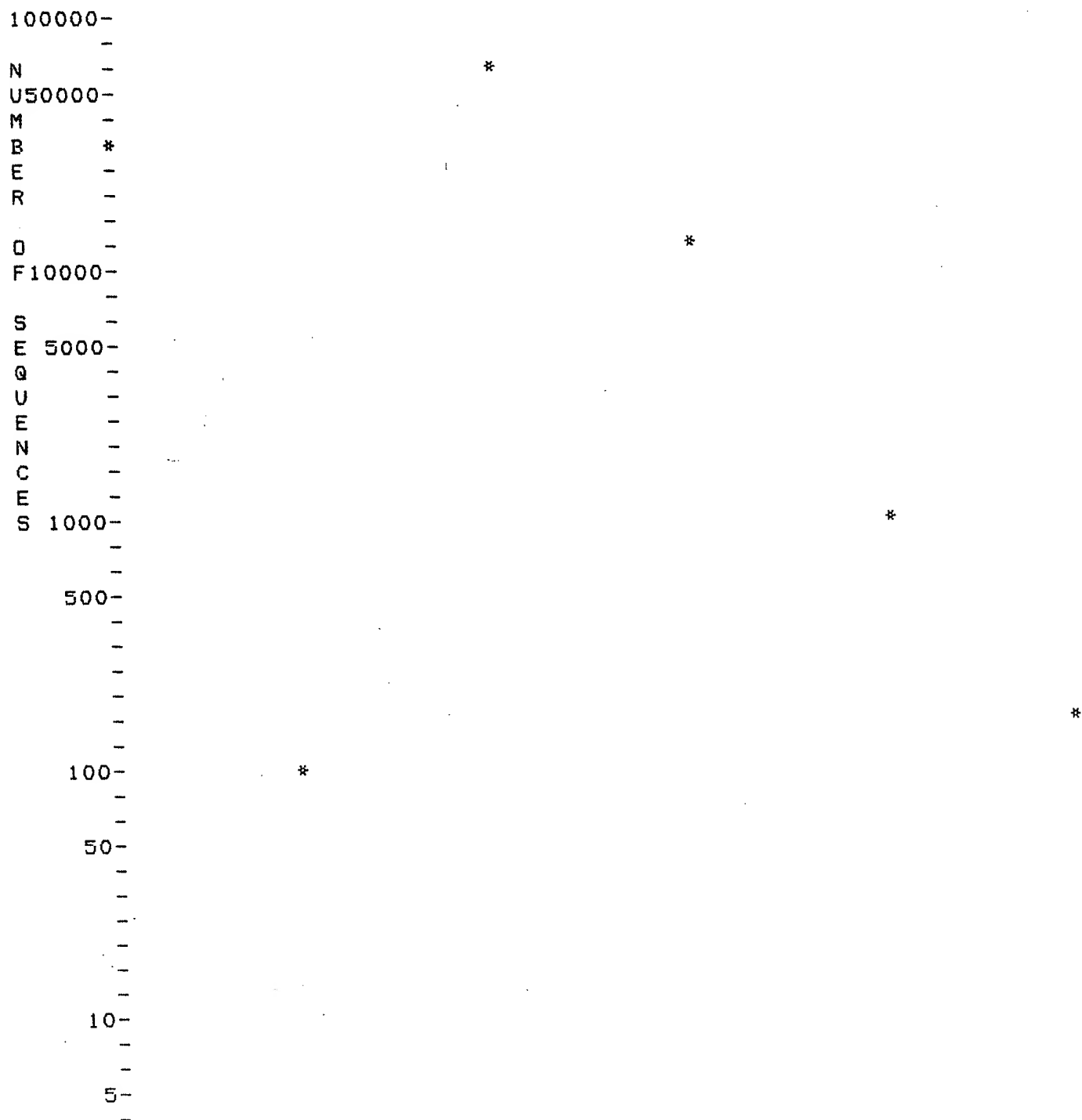
> 0 <
0| 10 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-14.res made by alexk on Thu 25 Feb 93 10:28:57-PST.

Query sequence being compared: CELSA-14 (1-5)
Number of sequences searched: 94553
Number of scores above cutoff: 4598

Results of the initial comparison of CELSA-14 (1-5) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

```

-----
0-----
||      |      ||      |      |      |      |      |      |
SCORE 0|      1      |1      2      | 2      3      | 3      4      4      5
STDEV  0      1      1      2      | 2      3      | 3      4      4      5

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5      Joining penalty      20
Gap penalty      1.00      Window size      5
Gap size penalty      0.26
Cutoff score      0
Randomization group      0

Initial scores to save      50      Alignments to save      0
Optimized scores to save      0      Display context      0

```

SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              1          3          1.05

```

```

Times:      CPU      Total Elapsed
            00:03:18.06      00:06:52.00

```

```

Number of residues:      25433612
Number of sequences searched:      94553
Number of scores above cutoff:      4598

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

145 100% similar sequences to the query sequence were found:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
1. S18268	*alpha-Aminoadipyl-L-cysteiny	3649	5	5	3.82	0
2. ACVS_NOCLA	L-(ALPHA-AMINOADIPYL)-L-CYSTE	3649	5	5	3.82	0
3. S12332	Ubiquitin--protein ligase - Y	1950	5	5	3.82	0
4. UBR1_YEAST	N-END-RECOGNIZING PROTEIN (UB	1950	5	5	3.82	0
5. RRWGNV	RNA-directed RNA polymerase -	1643	5	5	3.82	0
6. VDR1_NMV	186 KD PROTEIN (ORF 1).	1643	5	5	3.82	0
7. WMWGPV	RNA-directed RNA polymerase -	1456	5	5	3.82	0
8. S14005	*Hypothetical protein, 166K -	1456	5	5	3.82	0
9. VDR1_PVX	165 KD PROTEIN (ORF 1).	1456	5	5	3.82	0
10. VDR1_PVXCP	165 KD PROTEIN (ORF 1).	1456	5	5	3.82	0
11. VDR1_PVXX3	165 KD PROTEIN (ORF 1).	1456	5	5	3.82	0
12. VGIHJ2	E2 glycoprotein precursor - M	1376	5	5	3.82	0
13. VGL2_CVM4	E2 GLYCOPROTEIN PRECURSOR (SP	1376	5	5	3.82	0
14. VGIH59	E2 glycoprotein precursor - M	1324	5	5	3.82	0
15. VGL2_CVMA5	E2 GLYCOPROTEIN PRECURSOR (SP	1324	5	5	3.82	0
16. VGIHMJ	E2 glycoprotein precursor - M	1235	5	5	3.82	0
17. VGL2_CVMJH	E2 GLYCOPROTEIN PRECURSOR (SP	1235	5	5	3.82	0
18. A40986	*M-cadherin - Mouse (fragment	730	5	5	3.82	0
19. SAHHA4E	Cell surface antigen 4EP heav	529	5	5	3.82	0

too many to
print out complete
I've printed the
1st 20 ref

21. A3773Y	*Cell surface antigen 4F2 hea	529	5	5	3.82	0
22. 4F2_HUMAN	4F2 CELL-SURFACE ANTIGEN HEAV	529	5	5	3.82	0
23. S20477	Ribulose-bisphosphate carboxy	498	5	5	3.82	0
24. RKFPLP	Ribulose-bisphosphate carboxy	485	5	5	3.82	0
25. RKFPLB	Ribulose-bisphosphate carboxy	485	5	5	3.82	0
26. D34921	*Ribulose-bisphosphate carbox	485	5	5	3.82	0
27. RBL_FLABI	RIBULOSE BISPHOSPHATE CARBOXY	485	5	5	3.82	0
28. C34921	*Ribulose-bisphosphate carbox	485	5	5	3.82	0
29. RBL_FLAPR	RIBULOSE BISPHOSPHATE CARBOXY	485	5	5	3.82	0
30. RBL_PHYAM	RIBULOSE BISPHOSPHATE CARBOXY	482	5	5	3.82	0
31. RBL_STEHA	RIBULOSE BISPHOSPHATE CARBOXY	482	5	5	3.82	0
32. RBL_ALLPR	RIBULOSE BISPHOSPHATE CARBOXY	480	5	5	3.82	0
33. S20847	*Ribulose-bisphosphate carbox	480	5	5	3.82	0
34. RBL_BASAL	RIBULOSE BISPHOSPHATE CARBOXY	480	5	5	3.82	0
35. RBL_MOLVE	RIBULOSE BISPHOSPHATE CARBOXY	480	5	5	3.82	0
36. RBL_RIVHU	RIBULOSE BISPHOSPHATE CARBOXY	480	5	5	3.82	0
37. RBL_TRIPO	RIBULOSE BISPHOSPHATE CARBOXY	480	5	5	3.82	0
38. RKNULT	Ribulose-bisphosphate carboxy	478	5	5	3.82	0
39. G34921	*Ribulose-bisphosphate carbox	478	5	5	3.82	0
40. RBL_NEUMU	RIBULOSE BISPHOSPHATE CARBOXY	478	5	5	3.82	0
41. RBL_NEUTE	RIBULOSE BISPHOSPHATE CARBOXY	478	5	5	3.82	0
42. RKNULM	Ribulose-bisphosphate carboxy	478	5	5	3.82	0
43. H34921	*Ribulose-bisphosphate carbox	478	5	5	3.82	0
44. P50585	Maize ribulose-biphosphate ca	477	5	5	3.82	0
45. RKNTLO	Ribulose-bisphosphate carboxy	477	5	5	3.82	0
46. RKNTLA	Ribulose-bisphosphate carboxy	477	5	5	3.82	0
47. RKRZL	Ribulose-bisphosphate carboxy	477	5	5	3.82	0
48. RBL_NICAC	RIBULOSE BISPHOSPHATE CARBOXY	477	5	5	3.82	0
49. RBL_NICOT	RIBULOSE BISPHOSPHATE CARBOXY	477	5	5	3.82	0
50. RBL_ORYSA	RIBULOSE BISPHOSPHATE CARBOXY	477	5	5	3.82	0

No alignments saved.

Variable	Value
100000	-
N	-
U50000	*
M	-
B	-
E	*
R	-
	*
D	-
F10000	-
S	-
E 5000	*
Q	-
U	-
E	-
N	-
C	-
E	-
S 1000	-
	-
500	-
	-
	*
	-
100	-
	-
50	-
	-
	-
10	-
	-
5	-
	-
	*

SCORE	0	1	2	3	4	5	6	7
STDEV	-1	0	1	2	3	4	5	6

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		

Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.06

Times:	CPU	Total Elapsed
	00:03:32.04	00:07:11.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4356

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 4 standard deviations above mean ****						
1. DP3B_STRCD	DNA POLYMERASE III, BETA CHAI	376	7	7	4.71	0
**** 3 standard deviations above mean ****						
2. UL96_HCMVA	HYPOTHETICAL PROTEIN UL96.	115	6	6	3.77	0
3. S09861	Hypothetical protein UL96 - H	115	6	6	3.77	0
4. FKBP_NEUCR	FK506-BINDING PROTEIN (FKBP)	120	6	6	3.77	0
5. S11090	*FK506-binding protein - Neur	120	6	6	3.77	0
6. P00185	Tonoplast intrinsic protein b	169	6	6	3.77	0
7. POL_SFV1	POL POLYPROTEIN (REVERSE TRAN	193	6	7	3.77	0
8. A33562	*pol polyprotein - Simian foa	193	6	7	3.77	0
9. S05586	Hypothetical protein 2 - Maiz	216	6	6	3.77	0
10. A41378	*Hypothetical protein (cycB 5	221	6	6	3.77	0
11. A05182	Alcohol dehydrogenase beta -	243	6	6	3.77	0
12. A31087	Alzheimer's disease amyloid b	264	6	6	3.77	0
13. P90609	Sequence of amy 37 clone.	264	6	6	3.77	0
14. P90497	Protein sequence (includes th	274	6	6	3.77	0

16. KPR2_HUMAN	RIBOSE-PHOSPHATE PYROPHOSPHOK	317	6	6	3.77	0
17. KPR1_HUMAN	RIBOSE-PHOSPHATE PYROPHOSPHOK	318	6	6	3.77	0
18. KIRTR1	Ribose-phosphate pyrophosphok	318	6	6	3.77	0
19. KIHUR1	Ribose-phosphate pyrophosphok	318	6	6	3.77	0
20. KIRTR2	Ribose-phosphate pyrophosphok	318	6	6	3.77	0
21. KIHUR2	Ribose-phosphate pyrophosphok	323	6	6	3.77	0
22. COBD_PSEDE	COBD PROTEIN.	323	6	6	3.77	0
23. E36144	*cobD protein - Pseudomonas s	323	6	6	3.77	0
24. R13495	P.denitrificans COB D.	338	6	6	3.77	0
25. CELF_HSV2H	CELL FUSION PROTEIN PRECURSOR	338	6	6	3.77	0
26. JH0143	Syn protein - Human herpesvir	395	6	6	3.77	0
27. YT44_STRFR	HYPOTHETICAL 44.4 KD PROTEIN	395	6	6	3.77	0
28. JQ0430	Hypothetical 44.4K protein -	412	6	6	3.77	0
29. A30320	*Alzheimer's disease amyloid	412	6	6	3.77	0
30. A29030	Alzheimer's disease amyloid A	441	6	6	3.77	0
31. NCAP_DUGBV	NUCLEOCAPSID PROTEIN (NUCLEOP	441	6	6	3.77	0
32. VHVUDU	Nucleocapsid protein N - Dugb	484	6	6	3.77	0
33. A32761	*Alzheimer's disease amyloid	493	6	6	3.77	0
34. GAG_HIV1U	GAG POLYPROTEIN (CORE PROTEIN	500	6	6	3.77	0
35. GAG_HIV1N	GAG POLYPROTEIN (CORE PROTEIN	500	6	6	3.77	0
36. P91256	Recombinant gag precursor of	523	6	6	3.77	0
37. OCTC_RAT	CARNITINE OCTANOYLTRANSFERASE	523	6	6	3.77	0
38. A31948	Carnitine octanoyltransferase	574	6	6	3.77	0
39. B30320	*Alzheimer's disease amyloid	596	6	6	3.77	0
40. UL32_HSV11	HYPOTHETICAL PROTEIN UL32.	596	6	6	3.77	0
41. WMBEH2	UL32 protein - Human herpesvi	617	6	6	3.77	0
42. HEMA_MEASH	HEMAGGLUTININ-NEURAMINIDASE (617	6	6	3.77	0
43. HEMA_MEASE	HEMAGGLUTININ-NEURAMINIDASE (617	6	6	3.77	0
44. HMNZED	Hemagglutinin - Measles virus	617	6	6	3.77	0
45. HMNZHA	Hemagglutinin - Measles virus	617	6	6	3.77	0
46. HEMA_MEASY	HEMAGGLUTININ-NEURAMINIDASE (620	6	6	3.77	0
47. JU0273	Hemagglutinin - Measles virus	620	6	6	3.77	0
48. S18607	*Phosphotransferase system en	651	6	6	3.77	0
49. B41080	*Probable transketolase - Rho	657	6	6	3.77	0
50. A4_RAT	ALZHEIMER'S DISEASE AMYLOID A	695	6	6	3.77	0

No alignments saved.

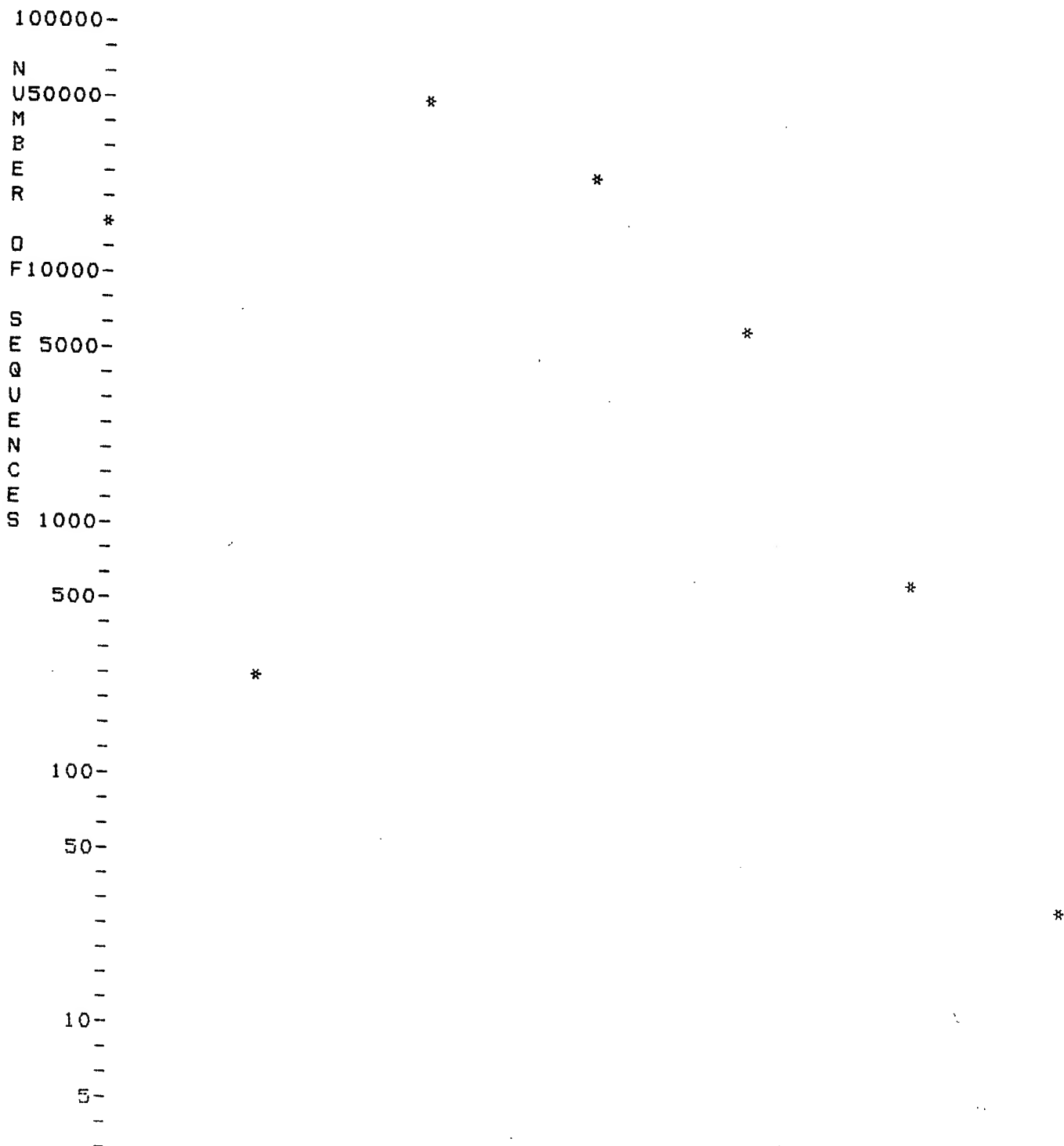
0| 0 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-16.res made by alexk on Thu 25 Feb 93 10:45:14-PST.

Query sequence being compared: CELSA-16 (1-11)
Number of sequences searched: 94553
Number of scores above cutoff: 4223

Results of the initial comparison of CELSA-16 (1-11) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

```

-
-
-
0-----
SCORE 0|      |      |      |      |      |      |      |      |      |
STDEV -1|      |      |      |      |      |      |      |      |      |
      1|      |      |      |      |      |      |      |      |      |
      0|      |      |      |      |      |      |      |      |      |
      1|      |      |      |      |      |      |      |      |      |
      2|      |      |      |      |      |      |      |      |      |
      1|      |      |      |      |      |      |      |      |      |
      3|      |      |      |      |      |      |      |      |      |
      1|      |      |      |      |      |      |      |      |      |
      3|      |      |      |      |      |      |      |      |      |
      4|      |      |      |      |      |      |      |      |      |
      3|      |      |      |      |      |      |      |      |      |
      5|      |      |      |      |      |      |      |      |      |
      5|      |      |      |      |      |      |      |      |      |
      6|      |      |      |      |      |      |      |      |      |

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5      Joining penalty      20
Gap penalty      1.00      Window size      5
Gap size penalty      0.26
Cutoff score      0
Randomization group      0

Initial scores to save      50      Alignments to save      0
Optimized scores to save      0      Display context      0

```

SEARCH STATISTICS

```

Scores:      Mean      Median      Standard Deviation
              2          3          1.08

```

```

Times:      CPU      Total Elapsed
            00:03:26.11      00:07:10.00

```

```

Number of residues:      25433612
Number of sequences searched:      94553
Number of scores above cutoff:      4223

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 3 standard deviations above mean ****						
1. PE25_NPVAC	25.1 KD PROTEIN IN PE-P26 INT	219	6	6	3.71	0
2. NTRB_VIBAL	NITROGEN REGULATION PROTEIN N	350	6	7	3.71	0
3. JL0114	ntrB protein - Vibrio alginol	350	6	7	3.71	0
4. GUN_CELUD	ENDOGLUCANASE PRECURSOR (EC 3	359	6	6	3.71	0
5. P70396	Cellulase.	359	6	6	3.71	0
6. IL1S_HUMAN	INTERLEUKIN-1 RECEPTOR, TYPE	398	6	6	3.71	0
7. S17428	*Interleukin-1 receptor, type	398	6	6	3.71	0
8. R15614	Human type II interleukin-1 r	398	6	6	3.71	0
9. R12786	Actinomyces Phospholipase D.	524	6	6	3.71	0
10. UL32_EBV	HYPOTHETICAL PROTEIN BFLF1.	525	6	6	3.71	0
11. Q0BE6	BFLF1 protein - Human herpesv	525	6	6	3.71	0
12. SPG7_DICDI	SPORE GERMINATION PROTEIN 270	532	6	6	3.71	0
13. B35621	Spore germination protein 270	532	6	6	3.71	0
14. VGLE_HSV11	GLYCOPROTEIN E PRECURSOR.	550	6	6	3.71	0
15. VGBE18	Glucoprotein E - Human herpesv	550	6	6	3.71	0

17.	S19202	*Glycoprotein Ben precursor -	588	6	7	3.71	0
18.	JH0506	Adhesion molecule SC1 precurs	588	6	7	3.71	0
19.	JH0593	Schwann cell myelin protein p	620	6	6	3.71	0
20.	VP3_RDV	MAJOR 114 KD STRUCTURAL PROTE	1019	6	6	3.71	0
21.	S12826	*Hypothetical protein - Rice	1019	6	6	3.71	0
22.	S12621	*114K major core protein - Ri	1019	6	6	3.71	0
23.	S20548	*P-glycoprotein homolog - Yea	1362	6	6	3.71	0
24.	KRE5_YEAST	KRE5 PROTEIN PRECURSOR.	1365	6	6	3.71	0
25.	BVBYK5	KRE5 protein precursor - Yeas	1365	6	6	3.71	0
26.	A39401	*Merozoite surface antigen 1	1726	6	6	3.71	0
27.	TEGU_VZVD	LARGE TEGUMENT PROTEIN.	2763	6	6	3.71	0
28.	WZBE22	Gene 22 protein - Human herpe	2763	6	6	3.71	0
**** 2 standard deviations above mean ****							
29.	R12065	Antigenic peptide #2269 D5/19	15	5	5	2.78	0
30.	S11617	Ribosomal protein HS17 - Halo	30	5	5	2.78	0
31.	S19945	Protein kinase crk3 - Mouse (35	5	5	2.78	0
32.	P91884	Antigenic Epstein-Barr virus	48	5	5	2.78	0
33.	S02383	Probable membrane antigen CL3	57	5	5	2.78	0
34.	C27606	Ig heavy chain V-a region (p2	58	5	5	2.78	0
35.	S22133	*Hypothetical protein - Esche	76	5	5	2.78	0
36.	KPSJ_HUMAN	PUTATIVE SERINE/THREONINE-PRO	101	5	5	2.78	0
37.	C26368	Protein-serine kinase (clone	101	5	5	2.78	0
38.	NRAM_IADA1	NEURAMINIDASE (EC 3.2.1.18) (103	5	6	2.78	0
39.	CYTA_RAT	CYSTATIN ALPHA (EPIDERMAL THI	103	5	5	2.78	0
40.	A00891	Sialidase - Influenza A virus	103	5	6	2.78	0
41.	UDRTS2	Cystatin alpha - Rat	103	5	5	2.78	0
42.	HM20_XENLA	HOMEOTIC PROTEIN NRL-20 (FRAG	105	5	5	2.78	0
43.	S12181	*X1-pou protein - African cla	105	5	5	2.78	0
44.	S21856	*Collagen alpha 1(X) chain -	106	5	5	2.78	0
45.	S15826	*Collagen - Human	106	5	5	2.78	0
46.	KV6K_MOUSE	IG KAPPA CHAIN V-VI REGION (N	108	5	5	2.78	0
47.	PRVA_RAJCL	PARVALBUMIN ALPHA.	109	5	5	2.78	0
48.	S11125	*Ig heavy chain V region - Mo	109	5	5	2.78	0
49.	PVRYC	Parvalbumin - Thornback ray	109	5	5	2.78	0
50.	P70092	Sequence encoded by (2'-5') o	110	5	6	2.78	0

No alignments saved.

Best Available Copy

	0	1	2	3	4	5	6	7	8
SCORE	0	1	2	3	4	5	6	7	8
STDEV	-1	0	1	2	3	4			

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.13

Times:	CPU	Total Elapsed
	00:03:37.03	00:07:11.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4785

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 5 standard deviations above mean ****						
1. LV4A_HUMAN	IG LAMBDA CHAIN V-IV REGION (106	8	8	5.32	0
2. L4HUBU	Ig lambda chain V-IV region (106	8	8	5.32	0
**** 4 standard deviations above mean ****						
3. HMN3_DROME	HOMEODOMAIN PROTEIN NK-3 (FRAGME	194	7	7	4.43	0
4. C33976	*NK-3 homeotic protein - Frui	194	7	7	4.43	0
5. ATP6_STRPU	ATP SYNTHASE A CHAIN (EC 3.6.	229	7	7	4.43	0
6. S01505	H+-transporting ATP synthase	229	7	7	4.43	0
7. TRPD_BACSU	ANTHRANILATE PHOSPHORIBOSYLTR	337	7	7	4.43	0
8. B22794	Anthranilate phosphoribosyltr	337	7	7	4.43	0
9. NPBS	Anthranilate phosphoribosyltr	337	7	7	4.43	0
10. TRPD_BACPU	ANTHRANILATE PHOSPHORIBOSYLTR	340	7	7	4.43	0
11. JH0099	trpD protein - Bacillus pumil	340	7	7	4.43	0
**** 3 standard deviations above mean ****						
12. A31093	Stage IV sporulation protein	43	6	6	3.55	0
	Stage IV sporulation protein - M	60	6	6	3.55	0

15.	V7K_BYDVP	*HOX 2.4 homeotic protein - M	62	6	6	3.55	0
16.	S00951	6.7 KD PROTEIN (ORF 5).	63	6	6	3.55	0
17.	S15534	Hypothetical protein, 6.7K -	63	6	6	3.55	0
18.	S15533	Homeotic protein Hox 3A - Hum	66	6	6	3.55	0
19.	S13785	Homeotic protein Hox 2D - Hum	66	6	6	3.55	0
20.	HM24_HUMAN	*Homeotic protein m31 - Mouse	69	6	6	3.55	0
21.	B37042	HOMEOBOX PROTEIN HOX-2.4 (HOX	70	6	6	3.55	0
22.	NU4M_ARTSX	*Homeotic protein Hox 2.4 - H	70	6	6	3.55	0
23.	S01211	NADH-UBIQUINONE OXIDOREDUCTAS	71	6	8	3.55	0
24.	HM24_CHICK	NADH dehydrogenase (ubiquinon	71	6	8	3.55	0
25.	S10884	HOMEOBOX PROTEIN HOX-2.4 (FR	74	6	6	3.55	0
26.	S14975	*Homeotic protein Hox 2.4 - C	74	6	6	3.55	0
27.	HM54_HUMAN	*Extensin class II - Tomato	82	6	6	3.55	0
28.	B32830	HOMEOBOX PROTEIN HOX-5.4 (HOX	95	6	6	3.55	0
29.	S05957	*Hox5.4 homeotic protein - Hu	95	6	6	3.55	0
30.	YENM_SALTY	Homeotic protein Hox 4E - Hum	95	6	6	3.55	0
31.	S15521	HYPOTHETICAL 10.4 KD PROTEIN	99	6	6	3.55	0
32.	S16177	*Homeotic protein Hox 4.3 - M	99	6	6	3.55	0
33.	HMR4_RAT	*Homeotic protein Hox 4.3 - M	99	6	6	3.55	0
34.	HMRA_RAT	HOMEOBOX PROTEIN R4 (FRAGMENT	104	6	6	3.55	0
35.	KV3J_HUMAN	HOMEOBOX PROTEIN R1A (FRAGMEN	114	6	6	3.55	0
36.	K3HUVH	IG KAPPA CHAIN PRECURSOR V-II	116	6	6	3.55	0
37.	R15106	Ig kappa chain precursor V-II	116	6	6	3.55	0
38.	NODN_RHIME	hCG/bLH chimera, D10	145	6	6	3.55	0
39.	S16566	NODULATION PROTEIN N.	160	6	6	3.55	0
40.	IT1B_PSOTE	*nolN protein - Rhizobium mel	160	6	6	3.55	0
41.	IT1A_PSOTE	TRYPSIN INHIBITOR 1B (WTI-1B)	172	6	6	3.55	0
42.	TIWDKB	TRYPSIN INHIBITOR 1A (WTI-1A)	172	6	6	3.55	0
43.	TIWDK	Trypsin inhibitor 1B (Kunitz)	172	6	6	3.55	0
44.	FRI3_RANCA	Trypsin inhibitor 1A (Kunitz)	172	6	6	3.55	0
45.	B27805	FERRITIN, LOWER SUBUNIT (FERR	173	6	7	3.55	0
46.	CRTZ_ERWUR	Ferritin chain L - Bullfrog	174	6	7	3.55	0
47.	F37802	CRTZ PROTEIN (EC -.-.-.-).	175	6	6	3.55	0
48.	R07468	*crtZ protein - Erwinia uredo	175	6	6	3.55	0
49.	IT2_PSOTE	Polypeptide with enzymatic ac	175	6	6	3.55	0
50.		TRYPSIN INHIBITOR 2 (WTI-2).	182	6	6	3.55	0

No alignments saved.

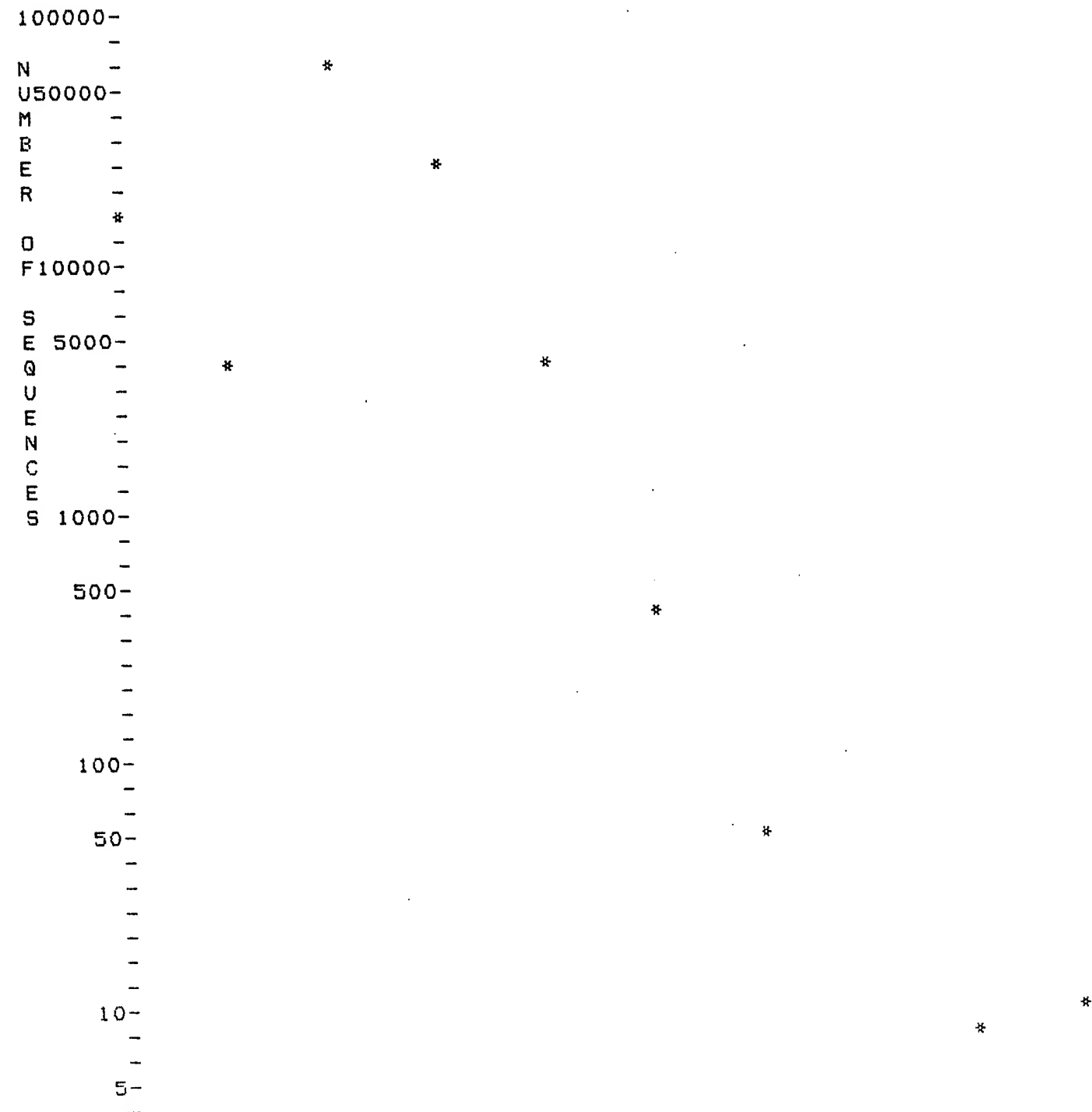
> 0 <
0| 0 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-18.res made by alexk on Thu 25 Feb 93 10:37:20-PST.

Query sequence being compared: CELSA-18 (1-15)
Number of sequences searched: 94553
Number of scores above cutoff: 3965

Results of the initial comparison of CELSA-18 (1-15) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

```

-
-
-
0-----
SCORE 0| 1| 2| 3| 4| 5| 6| 7| 8| 9
STDEV -1 0 1 2 3 4 5 6

```

PARAMETERS

```

Similarity matrix      Unitary      K-tuple      2
Mismatch penalty      5            Joining penalty 20
Gap penalty           1.00          Window size    5
Gap size penalty      0.26
Cutoff score          0
Randomization group    0

Initial scores to save 50            Alignments to save 0
Optimized scores to save 0          Display context    0

```

SEARCH STATISTICS

```

Scores:                Mean          Median      Standard Deviation
                        2              3           1.00

```

```

Times:                 CPU              Total Elapsed
                        00:03:29.06      00:07:15.00

```

```

Number of residues:      25433612
Number of sequences searched: 94553
Number of scores above cutoff: 3965

```

```

Cut-off raised to 2.
Cut-off raised to 3.
Cut-off raised to 4.
Cut-off raised to 5.

```

The scores below are sorted by initial score.
Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 6 standard deviations above mean ****						
1. PPD1_BOVIN	PHOSPHODIESTERASE I (EC 3.1.4	61	9	9	6.98	0
2. A25274	Phosphodiesterase I - Bovine	61	9	9	6.98	0
3. A41179	*Protein kinase PC-1 - Bovine	289	9	9	6.98	0
4. PC1_MOUSE	PLASMA-CELL MEMBRANE GLYCOPRO	871	9	9	6.98	0
5. PC1_HUMAN	PLASMA-CELL MEMBRANE GLYCOPRO	873	9	9	6.98	0
6. A27410	Plasma cell membrane protein	905	9	9	6.98	0
7. S21706	*Pyrophosphatase - Human	925	9	9	6.98	0
8. A39216	*Plasma cell membrane protein	925	9	9	6.98	0
**** 5 standard deviations above mean ****						
9. A37258	Hypothetical protein, 13K - M	115	8	8	5.98	0
10. A28336	Hypothetical protein, 13K - M	115	8	8	5.98	0
11. A27498	Coagulation Factor V precursor	1600	8	8	5.98	0
12. FAS_HUMAN	COAGULATION FACTOR V PRECURSOR	2224	8	8	5.98	0
13. A28028	Coagulation Factor V precursor	2224	8	8	5.98	0
14. TFCU_HUMAN	LARGE TGF- β RECEPTOR PROTEIN	2763	8	9	5.98	0

**** 3 standard deviations above mean ****

16. B32268	*Carcinoembryonic antigen pre	120	6	6	3.99	0
17. PQ0328	Mucin-like peptide - Human (f	141	6	6	3.99	0
18. SODC_HAEP	SUPEROXIDE DISMUTASE PRECURSO	187	6	6	3.99	0
19. SODC_HAEIN	SUPEROXIDE DISMUTASE LIKE PRO	187	6	6	3.99	0
20. B41654	*Superoxide dismutase (Cu-Zn)	187	6	6	3.99	0
21. A41654	*Superoxide dismutase (Cu-Zn)	187	6	6	3.99	0
22. P83190	[delta 1-5][Arg358] Alpha 1-a	390	6	6	3.99	0
23. P83189	[Ala357, Arg358] Alpha 1-anti	395	6	6	3.99	0
24. P40134	Sequence of human alpha-1-ant	418	6	6	3.99	0
25. P94665	Human alpha-1-antitrypsin as	418	6	6	3.99	0
26. FUCP_ECOLI	L-FUCOSE PERMEASE.	438	6	6	3.99	0
27. WQECFP	Fucose permease - Escherichia	438	6	6	3.99	0
28. GSH1_ECOLI	GLUTAMATE--CYSTEINE LIGASE (E	518	6	6	3.99	0
29. SYECEC	Glutamate--cysteine ligase -	518	6	6	3.99	0
30. TETO_CAMJE	TETRACYCLINE RESISTANCE PROTE	637	6	6	3.99	0
31. A29809	Tetracycline resistance prote	637	6	6	3.99	0
32. TETO_STRMU	TETRACYCLINE RESISTANCE PROTE	639	6	6	3.99	0
33. TETO_CAMCO	TETRACYCLINE RESISTANCE PROTE	639	6	6	3.99	0
34. S06388	*Tetracycline resistance prot	639	6	6	3.99	0
35. A31098	Tetracycline resistance prote	639	6	6	3.99	0
36. CR72_BACTI	72 KD CRYSTAL PROTEIN (DELTA	643	6	6	3.99	0
37. P91462	67-kD protein toxin.	643	6	6	3.99	0
38. DPOL_HPBV4	DNA POLYMERASE (EC 2.7.7.7).	730	6	6	3.99	0
39. DPOL_HPBVZ	DNA POLYMERASE (EC 2.7.7.7) (750	6	6	3.99	0
40. JDVLVH	DNA-directed DNA polymerase -	750	6	6	3.99	0
41. DPOM_HPBVY	DNA POLYMERASE (EC 2.7.7.7) (832	6	6	3.99	0
42. DPOL_HPBVY	DNA POLYMERASE (EC 2.7.7.7) (832	6	6	3.99	0
43. DPOL_HPBVA	DNA POLYMERASE (EC 2.7.7.7).	832	6	6	3.99	0
44. S20748	*Hypothetical protein (preS1,	832	6	6	3.99	0
45. S20757	*Hypothetical protein (pre S1	832	6	6	3.99	0
46. JDVLA1	DNA-directed DNA polymerase -	832	6	6	3.99	0
47. JDVLVB	DNA-directed DNA polymerase -	832	6	6	3.99	0
48. JDVLVA	DNA-directed DNA polymerase -	832	6	6	3.99	0
49. JDVLVS	DNA-directed DNA polymerase -	842	6	6	3.99	0
50. DPOL_HPBVR	DNA POLYMERASE (EC 2.7.7.7).	843	6	6	3.99	0

No alignments saved.

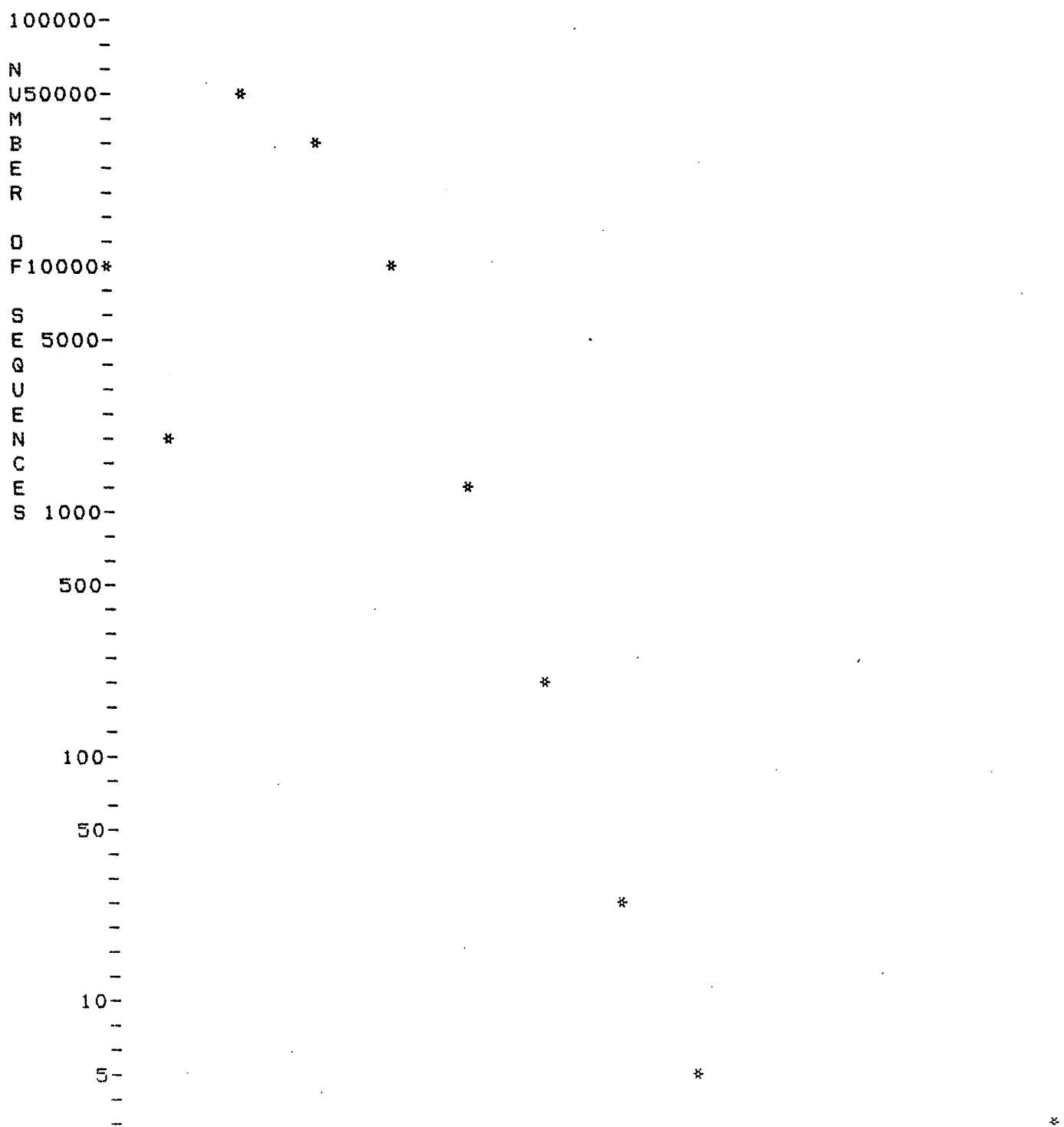
> 0 <
0| 0 IntelliGenetics
> 0 <

FastDB - Fast Pairwise Comparison of Sequences
Release 5.4

Results file celsa-19.res made by alexk on Thu 25 Feb 93 10:37:29-PST.

Query sequence being compared: CELSA-19 (1-22)
Number of sequences searched: 94553
Number of scores above cutoff: 4944

Results of the initial comparison of CELSA-19 (1-22) with:
Data bank : A-GeneSeq 9, all entries
Data bank : PIR 34, all entries
Data bank : Swiss-Prot 23, all entries



Best Available Copy

0-----

SCORE	0	1	2	3	4	5	6	7	8	9	10	11	12	13
STDEV	-1	0	1	2	3	4	5	6	7	8	9	10	11	12

PARAMETERS

Similarity matrix	Unitary	K-tuple	2
Mismatch penalty	5	Joining penalty	20
Gap penalty	1.00	Window size	5
Gap size penalty	0.26		
Cutoff score	0		
Randomization group	0		
Initial scores to save	50	Alignments to save	0
Optimized scores to save	0	Display context	0

SEARCH STATISTICS

Scores:	Mean	Median	Standard Deviation
	2	3	1.06

Times:	CPU	Total Elapsed
	00:03:41.96	00:07:24.00

Number of residues:	25433612
Number of sequences searched:	94553
Number of scores above cutoff:	4944

Cut-off raised to 2.
 Cut-off raised to 3.
 Cut-off raised to 4.
 Cut-off raised to 5.

The scores below are sorted by initial score.
 Significance is calculated based on initial score.

A 100% identical sequence to the query sequence was not found.

The list of best scores is:

Sequence Name	Description	Length	Init. Score	Opt. Score	Sig.	Frame
**** 10 standard deviations above mean ****						
1. PC1_HUMAN	PLASMA-CELL MEMBRANE GLYCOPRO	873	13	14	10.41	0
2. S21706	*Pyrophosphatase - Human	925	13	14	10.41	0
3. A39216	*Plasma cell membrane protein	925	13	14	10.41	0
**** 9 standard deviations above mean ****						
4. PC1_MOUSE	PLASMA-CELL MEMBRANE GLYCOPRO	871	12	13	9.46	0
5. A27410	Plasma cell membrane protein	905	12	13	9.46	0
**** 5 standard deviations above mean ****						
6. S20728	*Myosin light chain - Slime m	150	8	8	5.68	0
7. MLE_DICDI	MYOSIN, ESSENTIAL LIGHT CHAIN	166	8	8	5.68	0
8. A28127	Myosin light chain, nonmuscle	166	8	8	5.68	0
9. J00367	Hypothetical protein 332, plc	332	8	9	5.68	0
**** 4 standard deviations above mean ****						
10. PHA3_FREDI	C-PHYCOCYANIN-3 ALPHA CHAIN.	162	7	8	4.73	0
11. S05712	Phycocyanin 3 alpha chain - C	162	7	8	4.73	0
12. U435_UAPCC	PROTEIN A75	174	7	8	4.73	0

14. CY1_NEUCR	CYTOCHROME C1, HEME PROTEIN P	332	7	7	4.73	0
15. A27187	Cytochrome c1, heme protein p	332	7	7	4.73	0
16. JN0260	*Enterobacterial common antig	348	7	7	4.73	0
17. YIFC_ECOLI	HYPOTHETICAL 39.6 KD IN RFE 3	349	7	7	4.73	0
18. PAT1_SDLTU	PATATIN B1 PRECURSOR (POTATO	377	7	7	4.73	0
19. S05593	Patatin (clone pPATB1) - Pota	377	7	7	4.73	0
20. PAT5_SDLTU	PATATIN T5 PRECURSOR (POTATO	386	7	7	4.73	0
21. PAT0_SDLTU	PATATIN PRECURSOR (POTATO TUB	386	7	7	4.73	0
22. A26017	Patatin T5 precursor - Potato	386	7	7	4.73	0
23. JS0624	Fatty acid beta oxidation mul	391	7	7	4.73	0
24. VGL2_CVMJH	E2 GLYCOPROTEIN PRECURSOR (SP	1235	7	7	4.73	0
25. VGIHMJ	E2 glycoprotein precursor - M	1235	7	7	4.73	0
26. VGL2_CVMA5	E2 GLYCOPROTEIN PRECURSOR (SP	1324	7	7	4.73	0
27. VGIH59	E2 glycoprotein precursor - M	1324	7	7	4.73	0
28. VGL2_CVM4	E2 GLYCOPROTEIN PRECURSOR (SP	1376	7	7	4.73	0
29. VGIHJ2	E2 glycoprotein precursor - M	1376	7	7	4.73	0
**** 3 standard deviations above mean ****						
30. R13137	GP1b alpha peptide fragment.	15	6	7	3.79	0
31. THN6_HORVU	LEAF-SPECIFIC THIONIN (CLONE	46	6	6	3.79	0
32. S00825	Thionin BTH6, leaf - Barley	46	6	6	3.79	0
33. C551_ECTHA	CYTOCHROME C551.	78	6	7	3.79	0
34. CCER51	Cytochrome c551 - Ectothiorho	78	6	7	3.79	0
35. IMMC_ECOLI	IMMUNITY PROTEIN FOR CLOACIN.	85	6	6	3.79	0
36. B28585	Immunity protein - Escherichi	85	6	6	3.79	0
37. IMECP3	Immunity protein - Escherichi	85	6	6	3.79	0
38. FKBP_NEUCR	FK506-BINDING PROTEIN (FKBP)	120	6	6	3.79	0
39. S11090	*FK506-binding protein - Neur	120	6	6	3.79	0
40. THN1_HORVU	LEAF-SPECIFIC THIONIN PRECURS	137	6	7	3.79	0
41. THN4_HORVU	LEAF-SPECIFIC THIONIN PRECURS	137	6	7	3.79	0
42. S07648	Thionin precursor, leaf (clon	137	6	7	3.79	0
43. MLR_PHYPO	MYOSIN REGULATORY LIGHT CHAIN	147	6	7	3.79	0
44. S17769	*3-Dehydroquinase - Mycobacte	147	6	6	3.79	0
45. A29910	Myosin calcium-binding light	147	6	7	3.79	0
46. S09470	*Inter-alpha-trypsin inhibito	151	6	8	3.79	0
47. PHEA_MASLA	PHYCOCYANIN ALPHA CHAI	162	6	7	3.79	0
48. PHCA_SYNPW	C-PHYCOCYANIN ALPHA CHAIN.	162	6	7	3.79	0
49. PHA2_FREDI	C-PHYCOCYANIN-2 ALPHA CHAIN.	162	6	7	3.79	0
50. J00764	*Phycoerythrocyanin alpha cha	162	6	7	3.79	0

No alignments saved.

d 3

3. 5,003,047, Mar. 26, 1991, Method for purifying biologically active
ligate; Martin L. Yarmush, et al., 530/413, 412, 421, 427 [IMAGE
AVAILABLE]

=> log y

U.S. Patent & Trademark Office LOGOFF AT 14:10:12 ON 26 FEB 93

TABLE 4. Peptide sequences for Autotaxin.

	PEPTIDE NO.	AMINO ACID SEQUENCE	SEQ ID NO:	NAME
5	1.	WHVAAN	SEQ ID NO:1	ATX 18
	2.	PXLDVYK	SEQ ID NO:2	ATX 19
	3.	YPAFK	SEQ ID NO:3	ATX 20
	4.	QAEVS	SEQ ID NO:4	ATX 24
	5.	PEEVTXPNYL	SEQ ID NO:5	ATX 29
10	6.	YDVPWNETI	SEQ ID NO:6	ATX 47
	7.	SPPFENINLY	SEQ ID NO:7	ATX 48
	8.	GGQPLWITATK	SEQ ID NO:8	ATX 100
	9.	VNSMQTVFVGY-	SEQ ID NO:9	ATX 101
		GPTFK		
15	10.	DIEHLTSLDFFR	SEQ ID NO:10	ATX 102
	11.	TEFLSNYLTNVDD-	SEQ ID NO:11	ATX 103
		ITLVPGLGR		
	12.	QYLHQYGSS	SEQ ID NO:26	ATX 37
	13.	VLNYF	SEQ ID NO:27	ATX 39
20	14.	YLNAT	SEQ ID NO:28	ATX 40
	15.	HLLYGRPAVLY	SEQ ID NO:29	ATX 41
	16.	SYPEILTPADN	SEQ ID NO:30	ATX 44
	17.	XYGFLFPPYLSSSP	SEQ ID NO:31	ATX 53
	18.	TFPNLYVF/LAQGLYWS	SEQ ID NO:32	ATX 59
25	19.	VNVISGPIDYDYGDLH/A-	SEQ ID NO:33	ATX 104
		DTEDK		

Peptide numbers 1-7 refer to peaks numbered in FIG. 11. Peptide numbers 12-18 refer to peptides purified from the preparation which yielded peptide numbers 1-7. Peptides 8-11 and 19, are from a separate purification, not shown in FIG. 11.

X refers to potentially glycosylated residues.

d his

(FILE 'HOME' ENTERED AT 13:34:23 ON 26 FEB 93)

FILE 'REGISTRY' ENTERED AT 13:34:37 ON 26 FEB 93

L1 Ø S DIEHLTSLDFFR/SQEP
L2 Ø S DIEHLTSLDFFR/SQSP

=> log y

COST IN U.S. DOLLARS

SINCE FILE
ENTRY
26.20

TOTAL
SESSION
26.46

FULL ESTIMATED COST

STN INTERNATIONAL LOGOFF AT 13:36:24 ON 26 FEB 93